FINAL SYNOPSIS

Identification of novel target genes in burrowing nematode (*Radopholus similis* Thorne) through genome wide transcriptome analysis

Submitted by

Blessy M. Baby

Ph.D. Scholar,
ICAR-Indian Institute of Spices Research, Kozhikode

Guide

Dr. Santhosh J Eapen

Former Head &Principal Scientist, Crop Protection,
ICAR-Indian Institute of Spices Research, Kozhikode.

Submitted to
Mangalore University
Mangalagangothiri, Karnataka, India

ICAR-INDIAN INSTITUTE OF SPICES RESEARCH
(*Indian Council of Agricultural Research*)
PB No. 1701, Marikunnu P.O.
Kozhikode- 673012, Kerala
India

FINAL SYNOPSIS

Synopsis to the PhD work entitled "Identification of novel target genes in burrowing nematode (*Radopholus similis* Thorne) through genome wide transcriptome analysis" is as given below.

INTRODUCTION

Radopholus similis, commonly known as burrowing nematode, is a migratory endoparasitic nematode of many agricultural crops. *R. similis* (Nematoda, Tylenchida, Pratylenchidae) was first observed by Nathan A. Cobb in necrotic banana roots from Fiji in 1891. It is a notorious pathogen, causing the yellow disease or slow decline (wilt) of black pepper, spreading decline of citrus and toppling disease of banana. The burrowing nematode is an obligate parasite and needs a living host to survive, though its various stages can move from root to soil and vice versa. *Radopholus* spp. are considered to be among the 10 most damaging plant parasitic nematodes (PPN) world-wide (Sasser and Freckman, 1987; EPPO Diagnostics Report, 2008). As a result of the devastating effects *R. similis* can have on agricultural crops, EPPO phytosanitary categorization has classified the pathotype that can attack banana as an A2 quarantine organism, and the pathotype that can attack citrus as A1. These nematodes have caused large losses to the citrus industry and completely destroyed black pepper plantations in Indonesia (MacGowan, 1977). *R. similis* continues to be a major pest problem for the banana industry in many parts of the world, and it has also been introduced into heated glasshouses in Europe by way of trade of ornamental plants, in particular Anthurium (EPPO Diagnostics Report, 2008).

Once burrowing nematode has become established in a field, the management of this nematode is difficult and complex. There is no known control method for complete and successful eradication of this nematodes from crop fields. Therefore, most of the current management strategies aim atreducing their preplant density to non-injurious levels. Since none of the classical control measures are efficient for theircontrol, a combination of various methods is recommended for their management. Some of the most significant management strategies currently being used involve the application of chemical pesticides in one form or the other. Though nematicides are generally effective in controlling nematodes and are easy to use, they are expensive, highly toxic and have a negative impact on the environment and human health. Complete reliance on chemical nematicides is no more advisable as they are proven to be carcinogenic and toxic to non-target organisms. Therefore, it is imperative to explore novel approaches to develop and establish effective and environmentally friendly nematode control methods.

Next generation sequencing (NGS) has greatly increased our ability to analyze transcriptomes and genomes of many organisms in a cost-effective manner (Wang *et al.,* 2009). Whole transcriptome RNA-sequencing (RNA-seq) is becoming increasingly attractive as a new approach for quantitative studies of genes and genomes of any species. Unparalleled access to sequence and expression variation is obtained in this method and it also allows for additional insights into molecular mechanisms like alternative splicing, cis vs. trans gene regulation or small non-coding RNA dynamics in the species of interest. RNA-seq studies prove to be more challenging in non-model species than in model organisms due to lack of quality genome assemblies (Hekman *et al.,* 2015). Nonetheless, there are increasingly more studies being focused on PPN using RNA-seq approach (Haegeman *et al.,* 2013; Eves-van den

Akker *et al.,* 2014; Bauters *et al.,* 2014; Elashry *et al.,* 2020; Petitot *et al.,* 2020) mostly devoted to sedentary plant-parasitic nematodes. However, of late, such studies are increasingly being conducted in migratory nematode species (Haegeman *et al.,* 2011; Nicol *et al.,* 2012; Vieira *et al.,* 2015, 2020; Huang *et al.,* 2019; Marmonier *et al.,* 2019; Channale *et al.,* 2021).

Reliable identification of PPNs is a prerequisite for effective management of nematodesand for understanding their genetic variation and evolution. Nematodes, in general, are difficult to identify because of their microscopic nature and lack of distinguishable morphological characters warranting use of DNA-based molecular tools for accuracy and reliability. The two most commonly used repeated regions for taxonomic and diagnostic purposes in nematodes are regions of ribosomal DNA (rDNA) and mitochondrial DNA (mtDNA). The emergence of rapid and inexpensive sequencing technologies has allowed researchers to analyze large multi-locus, even genomic-scale, datasets for phylogenetic analysis across the tree of life. Many different DNA-based methods have been reported for the identification and characterization of *Radopholus* spp. (Kaplan *et al.,* 2000; Elbadri *et al.,* 2002; Uribe *et al.,* 2010) but all these studies have been done using one or more genes, mostly ribosomal genes. Although mtDNA have recently been widely used in molecular characterization of PPNs (Blouin, 2002; Madani *et al.,* 2010; Handoo *et al.,* 2016; Bai *et al.,* 2020; Xu *et al.,* 2020) but it has not yet been utilized for the characterization of *R. similis*. Therefore, in this study two mitochondrial genes - cytochrome c oxidase gene (*cox1*) and NADH dehydrogenase gene (nd1), and four rDNA genes - the internal transcribed spacer region (ITS1-5.8S-ITS2),D3 expansion domain segments of the 28S gene, 18S rRNA and 28S rRNA genes - were concatenated and used to construct the phylogeny of *R. similis* populations across the world. For the first time these genes from an Indian isolate were amplified, cloned, sequenced and compared with the rest of the populations.

Secondly the study focused on transcriptome sequencing for the identification of Excretory/Secretory proteins in *R. similis*. Like in other plant parasitic nematodes, *R. similis* produce an array of secretory proteins that play an important role in migration of the nematode through root tissues and also in withdrawing the nutrients from the host cytoplasm through feeding (Davis *et al.,* 2000). Due to their importance in the life cycle of PPNs, a great deal of effort has been taken by several workers to identify various effectors using genomic, transcriptomic and proteomic analyses (Bellafiore*et al.,* 2008; Haegeman*et al.,* 2013; Maier *et al.,* 2013; Bauters*et al.,* 2014; Cotton *et al.,* 2014; Thorpe *et al.,* 2014; Petitot*et al.,* 2016; Gahoi*et al.,* 2019; Vieira *et al.,* 2020; Da Rocha *et al.,* 2021; Li *et al.,* 2021).Genes involved in the Excretory/Secretory pathways play crucial roles in the regulation of many other essential genes and cellular processes. The protein products of most of these genes have multiple conserved domains, which are involved in various functions implying that they may be involved in several disparate cellular processes, making them critical to the survival of an organism. Silencing of any such gene may have an effect on several mechanisms and for this reason, may severely impair development and viability of the organism (Iqbal *et al.,* 2020).

Urwin et al.,(2002), first investigated gene functions of plant parasitic nematodes, *Globodera pallida* and *Heterodera glycines* using *in vitro* RNAi (*in vitro* soaking with dsRNA). Subsequently, the same method has been successfully applied for examining the functions of genes in plant nematodes and other organisms. For plant parasitic nematodes, *in vitro* RNAi (performed by *in vitro*soaking of the nematodes

in a solution of dsRNA) is the most widely used method (Rosso *et al.,* 2005; Chen *et al.,* 2005; Shingles *et al.,* 2007; Li *et al.,* 2011, 2015; Wang *et al.,* 2012; Zhang*et al.,* 2012; Tan *et al.,* 2013; Blyuss *et al.,* 2019; Tian *et al.,* 2019; Iqbal *et al.,* 2020; Mani*et al.,* 2020). The major aim of this study was to evaluate if knockdown of seven genes which are important ES proteins is possible, and how such disruption, if any, will affect the behavior and infectivity of *R. similis*. This study proposes a catalog of RNAi phenotypes of these gene that may play an important role in nematode control via host-induced gene silencing and for deep learning of the mechanism of RNAi of *R. similis* and PPNs, of which very little is known at present.

OBJECTIVES

- Mining of genomic resources, assembly and annotation of putative genes of *R. similis*
- Identification and sequencing of novel genes in *R. similis* that can be used as targets for nematode control
- Experimental validation of the predicted novel target genes in *R. similis*
- Developing a bioinformatics pipeline for target gene identification in nematodes

REVIEW OF LITERATURE

ROLE OF EXCRETORY/SECRETOY PROTEINS IN PPNs

The protein molecules secreted by a cell is usually referred to as Excretory/Secretory (ES) proteins or secretome of that organism and includes various types of molecules like antibodies, antimicrobial peptides, digestive enzymes, extracellular proteinases, hormones, neurotransmitters and toxins. These proteins known as effectors, have various functions in the plant cell. The function of many of these effectors is less clear or unknown (Hageman *et al.,* 2011). ES proteins are involved across a wide range of taxa, in various essential biological processes including acquirement of nutrients, cell-cell communication, cell adhesion and migration, proteolysis, morphogenesis and regulation of immune responses (Maizels and Yazdanbakhsh, 2003; Cho *et al.,* 2002; Cho *et al.,* 2004). In case of plant parasitic nematodes, the candidate parasitism proteins secreted by nematodes to modify plant tissues includes cell-wall-modifying enzymes, multiple regulators of host cell cycle and metabolism, proteins that can localize to the plant cell nucleus, potential suppressors of host defense, mimics of plant molecules, and a relatively large cadre of predicted novel nematode parasitism proteins (Davis *et al.,* 2000).These effectors are believed to have been acquired from bacteria or fungi by horizontal gene transfer (Hageman *et al.,* 2011). The knowledge of ES pathways has facilitated reasonable progress in development of novel antibiotics in bacteria(Bonin-Debs *et al.,* 2004). Similarly, nematode secretions are believed to play key role in plant parasitism, making ES proteins an attractive target(Bonin-Debs *et al.,* 2004; Garg and Ranganathan, 2012) for the development of novel therapeutic solutions for parasitic infections. Various studies like proteomic approach, mass spectrometry technique or methods using

monoclonal antibodies (MAbs)are known for the identification and characterization of ES proteins in PPNs (Shinya *et al.,* 2013;Bellafiore, *et al.,* 2008; Smant, *et al.,* 1998; Rehman *et al.,* 2009; Rehman, *et al.,* 2016; Atkinson and Harris, 1989; Curtis, 2007; Robinson and Connolly, 2005). However, due to difficulty in the experimental functional allocation to individual proteins, the exact role of ES proteins in mediating various cellular processes is clandestine (Gahoi *et al.,* 2019). In view of this, computational approaches complemented with the experimental studies could be applied to identify and annotate encoded ES proteins. Numerous studies identified ES proteins using EST sequences(Garg and Ranganathan, 2012; Nagaraj *et al.,* 2008(a); Nagaraj *et al.,* 2008(b); Harcus *et al.,* 2004). With the aid of NGS technologies, several nematode genomes have been sequenced and the identification of ES proteins from sequenced genome used to prioritize the experimental investigation of new potential targets for parasitic diseases. The rapid increase of published sequence information of nematodes provides a new opportunity to systematically screen their ES proteins using bioinformatics methods.

RNAi BASED NEMATODE CONTROL

RNAi is a means by which sequence-specific post-transcriptional gene silencing is induced by dsRNA (double stranded RNA) (Fire *et al.,* 1998). First demonstrated in *C. elegans*, this method has subsequently been used in different organisms(Wang *et al.,*2016). RNAi is also a very powerful tool for examining the functions of genes in plant nematodes and other organisms. In case of PPNs, *in vitro* RNAi (performed by soaking the nematodes in a solution of dsRNA *in vitro*) is the most commonly used method (Wang *et al.,*2016). For different economically important species of PPNs (*Meloidogyne*, *Heterodera*, *Globodera*, *Pratylenchus*, *Bursaphelenchus*, and *Radopholus*), the function of particular genes in specific molecular or biological processes, have been explored using these approaches (Table 1). Different lengths and forms of dsRNA triggers have been used to accomplish gene knockdown in various life stages of these nematodes; these include short (e.g., 42 nucleotides, nt) and long dsRNAs (>100 nt), and 21 nt siRNAs (Dalzell *et al.,* 2010; Arguel *et al.,* 2012; Marmonier *et al.,* 2019; Iqbal *et al.,* 2020).

Table 1: List of genes silenced by RNAi in pant parasitic nematodes

Target gene	Nematode species	Phenotype	References
Hg-rps-23, BF014259	*H.glycines*	Decrease in J2 viability	Alkharouf *et al.,* 2007
Mi-ser-1, serine protease	*M.incognita*	Reduction in number of eggs per gram of root. Reduction in egg hatching ratio	Antonio de Souza Junior *et al.,* 2013
Mi-cpl-1, Cysteine protease	*M.incognita*	Reduction in number of eggs per gram of root.	Antonio de Souza Junior *et al.,* 2013
hg-pel	*H.glycines*	Favours male development	Bakhetia *et al.,* 2007
hg-eng-l , AF006052	*H.glycines*	Decrease in no. of established nematodes	Bakhetia *et al.,* 2007
hg-gp	*H.glycines*	Favours male development	Bakhetia *et al.,* 2007
hg-cm	*H.glycines*	Favours male development	Bakhetia *et al.,* 2007
hg-syv46, AF273728	*H.glycines*	Decrease in no. of established nematodes	Bakhetia *et al.,* 2007
pMiDuoxl, DQ082753	*M.incognita*	Upto 70% decrease in no of established nematodes Decrease in fecundity	Bakhetia *et al.,* 2007
Mispc3, Miduox	*M.incognita*	Reduction of nematode number root, retarded female development	Charlton *et al.,* 2010
Gr-ams-1, AJ270995	*G. rostochiensis*	Reduced ability to locate and invade roots	Chen *et al.,* 2005
Gr-eng-L, AF004523 Gr-eng-3, Gr-eng-4	*G. rostochiensis*	Reduced no. of established of established nematodes	Chen *et al.,* 2005; Rehman *et al.,* 2009

AF531170, parasitism gene	*M.incognita*	54-59% reduction in number of developing females	Choudhary *et al.,* 2012
Mc16D10L	*M.chitwoodi*	57 and 67% reduction in number of egg masses and eggs respectively	Dinh *et al.,* 2014
Mi-gsts-l , EL784458	*M.incognita*	52 %–71 % decreased in fecundity	Dubreuil *et al.,* 2007
Mi-cpl-1	*M.incognita*	60-80% reduction in infection and multiplication	Dutta *et al.,* 2015
Effector gene, Gp-hyp	*G. pallida*	Reduction in nematode parasitism	Eves-van den *et al.,* 2014
MjTIS-11	*M.javanica*		Fairbairn *et al.,* 2007
Unc-15	*Ditylenchus destructor*	50% reduction in the infection area	Fan *et al.,* 2015
AY013285	*M.artiellia*	Delayed egg hatch	Fanelli *et al.,* 2005
Mi-Cg-l	*M.incognita*	Avirulence gene being recognized by Mi-1 resistance gene	Gleason *et al.,* 2008
16D10 (DQ841121-DQ841123)	*M.arenaria, M. incognita, M. javanica, M. hapla,*	63-90% reduction no of galls and gall size	Huang *et al.,* 2006; Yang *et al.,* 2013
Fatty acid and retinol binding protein (Mj-far-1)	*M. javanica*	Ceased development of nematodes along with reduction in giant cell number	Iberkleid *et al.,* 2013
Tyrosine Phosphatase, Mitochondrial stress-70 protein precursors, Lactate dehydrogenase	*M.incognita*	Reduced no. of established females	Ibrahim *et al.,* 2011
Calreticulin-Mi-CRTN	*M.incognita*	Reduction in gall number	Jaouannet *et al.,* 2013
Gp-ftp-6	*G.pallida*	Inhibition of motility	Kimber *et al.,* 2007
Hg-rps-3a, CB379877	*H.glycines*	87 % reduction in female cysts	Klink *et al.,* 2009
Hg-rps-4, CB278739	*H.glycines*	81 % reduction in female cysts	Klink *et al.,* 2009
Hg-spk-1, BI451523.1	*H.glycines*	88 % reduction in female cysts	Klink *et al.,* 2009
Hg-snb-1, BF014436	*H.glycines*	93 % reduction in female cysts	Klink *et al.,* 2009
Y25, CB824330	*H. glycines*	81 % reduction in nematode eggs	Li *et al.,* 2010
Prp-17, AF113915	*H. glycines*	79 % reduction in nematode eggs	Li *et al.,* 2010
Cpn-1 , GU074018	*H.glycines*	95 % reduction in nematode eggs	Li *et al.,* 2010
Rs-cb-1	*R.similis*	Reduced reproduction and pathogenicity	Li *et al.,* 2015
Hg-amp-L, AY 883023	*H.glycines*	61 % decrease in number of female reproductive	Lilley *et al.,* 2005
HSP90, Heat Shock Protein	*M.incognita*	Delayed gall formation and upto 46% reduction in the number of eggs	Lourenco-Tessutti *et al.,* 2015
ICL, Isocitrate lyase	*M.incognita*	Upto 77% reduction in egg oviposition	Lourenco-Tessutti *et al.,* 2015
Mi-Rpn7	*M. incognita*	Reduction in reproduction and motility	Niu *et al.,* 2012
MiMAP40	*M.incognita*	Reduction in the number of galls	Niu *et al.,* 2016
flip-14, FMRF amide like peptide	*M.incognita*	Reduction in parasitic ability from 67-86%	Papolu *et al.,* 2013
flp-18, FMRF amide like peptide	*M.incognita*	Reduction in parasitic ability from 53-82%	Papolu *et al.,* 2013
Mi-crt, AF402771	*M.incognita*	Not determined	Rosso *et al.,* 2005
Mi-pg-1, AY098646	*M.incognita*	Not determined	Rosso *et al.,* 2005
Mi-cpl-l	*M.incognita*	60 % decrease in no. of established nematodes	Shingles *et al.,* 2007
4G06, AF469060	*H.schachtii*	23 %-64 % reduction in developing females	Sindhu *et al.,* 2009
3B05, AF469058	*H.schachtii*	12 %-47 % reduction in developing females	Sindhu *et al.,* 2009
8H07, AF500024	*H.schachtii*	>50 % reduction in developing females	Sindhu*et al.,* 2009
10A06, AF502391	*H.schachtii*	42 % reduction in developing females	Sindhu*et al.,* 2009
Major sperm protein	*H.glycines*	upto 68% reduction in number of eggs	Steeves *et al.,* 2006
hactl AF498244	*H.glycines*	41% decrese in no of established nematodes	Urwin *et al.,* 1997
hgcp-l	*H.glycines*	40% decrease in no of established nematodes	Urwin *et al.,* 1997
gpcp-i	*G.pallida*	25% less females ecovered	Urwin *et al.,* 1997
Pp-pat-10	*P.penetrans*	Upto 40% reduction in number of nematodes	Viera *et al.,* 2015
Pp-unc-87	*P.penetrans*	Upto 50% reduction in number of nematodes	Viera *et al.,* 2015

Pv010	*P. vulnus*	Reduced nematode multiplication with no visible lesions	Walawage *et al.,* 2013
8D05, Parasitism gene	*M.incognita*	Reduction in number of gals	Xue *et al.,* 2013
Integrase	*M.incognita*	>90% reduction in number of established nematodes	Yadav *et al.,* 2006
Splicing factor	*M.incognita*	>90% reduction in number of established nematodes	Yadav *et al.,* 2006
MeTCTP	*M.enterolobii*	Reduction in number of nematodes	Zhuo *et al.,* 2017
Rs-cps EU659125	*R. similis*	expression levels in nematodes decreased	Wang *et al.,* 2016
Rs-scp-1 KJ617041	*R. similis*	expression levels as well as pathogenicity in anthurium was significantly reduced	Huang *et al.,* 2019
RsVAP	*R. similis*	the symptoms of browning and necrosis and the degree of damage were relatively mild in the tomato roots	Li*et al.,* 2021
ego-1 and mes-2,	*M. incognita*	no significant ($p < 0.05$) change in expressions	Iqbal *et al.,* 2020
xpo-1, pash-1, xpo-2, rha-1, ekl-4, and csr-1	*M. incognita*	Two to nine-fold increase in gene expression	Iqbal *et al.,* 2020
rsd-3, xpo-2, drh-1, vig-1, eri-1, smg-2, smg-6, ppw-2, gfl-1, mut-7, mes-6 and 2242	*M. incognita*	Significantly reduced gene expression	Iqbal *et al.,* 2020

MATERIALS AND METHODS

1. MOLECULAR CHARACTERIZATION OF *R. SIMILIS*

1.1. R. similis isolation, culturing and maintenance

R. similis was cultured in carrots disc agar plates, according to the protocols obtained from O'Bannon and Taylor (1968) and David and Eric (1990)(Figure 1).

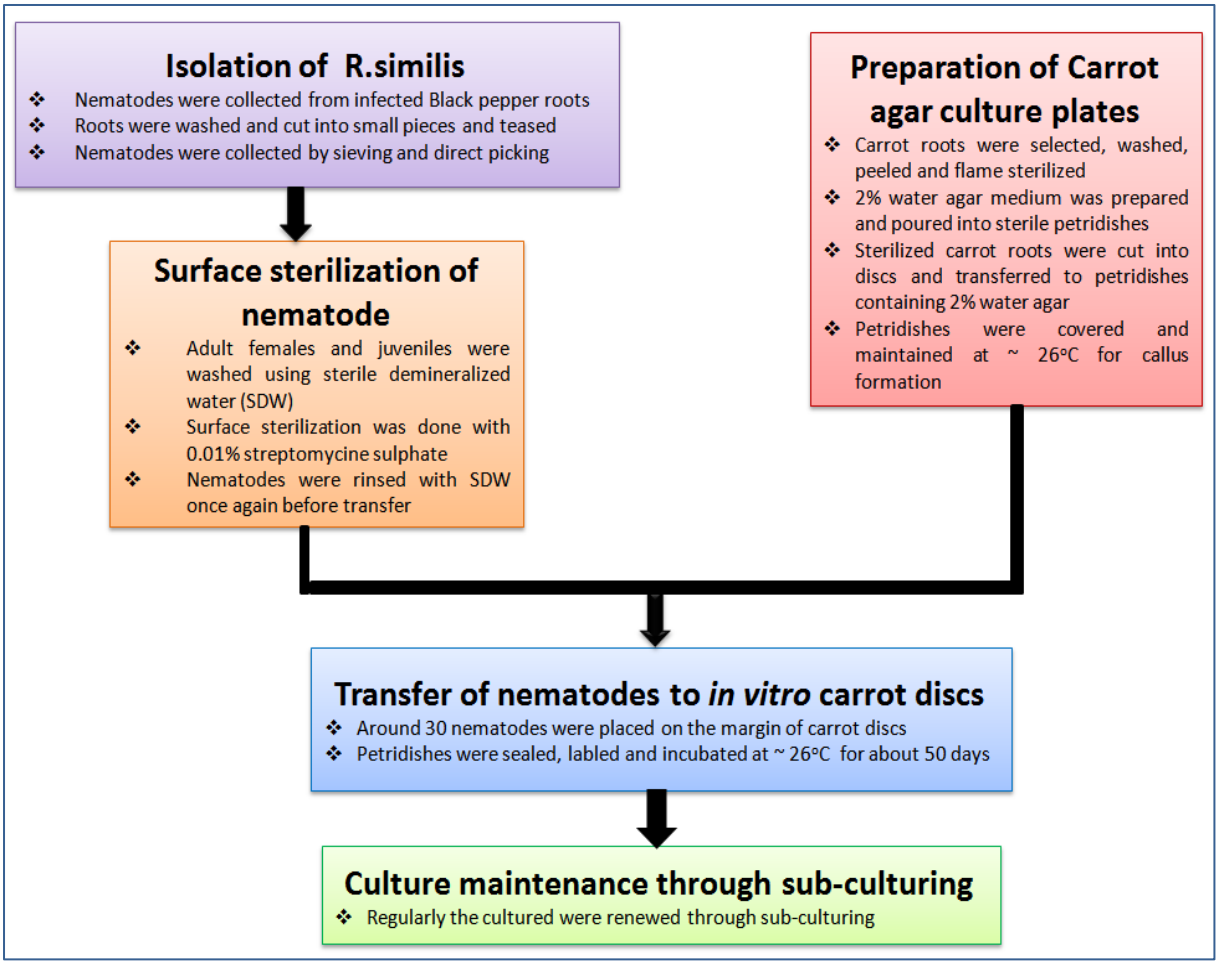

Isolation of R.similis
Nematodes were collected from infected Black pepper roots
Roots were washed and cut into small pieces and teased
Nematodes were collected by sieving and direct picking
Preparation of Carrot agar culture plates
Carrot roots were selected, washed, peeled and flame sterilized
2% water agar medium was prepared and poured into sterile petridishes
Sterilized carrot roots were cut into discs and transferred to petridishes containing 2% water agar
Petridishes were covered and maintained at ~ 26°C for callus formation
Surface sterilization of nematode
Adult females and juveniles were washed using sterile demineralized water (SDW)
Surface sterilization was done with 0.01% streptomycine sulphate
Nematodes were rinsed with SDW once again before transfer
Transfer of nematodes to in vitro carrot discs
Around 30 nematodes were placed on the margin of carrot discs
Petridishes were sealed, labled and incubated at ~ 26°C for about 50 days
Culture maintenance through sub-culturing
Regularly the cultured were renewed through sub-culturing

Figure 1. *R. similis* isolation and culturing methodology

Pictures of carrot disk agar plate with callus formation after 2 weeks, carrot disk culture plates and *R. similis* isolate R01(from black pepper)population obtained after 2 months, under stereo microscope are given in Figure 2.

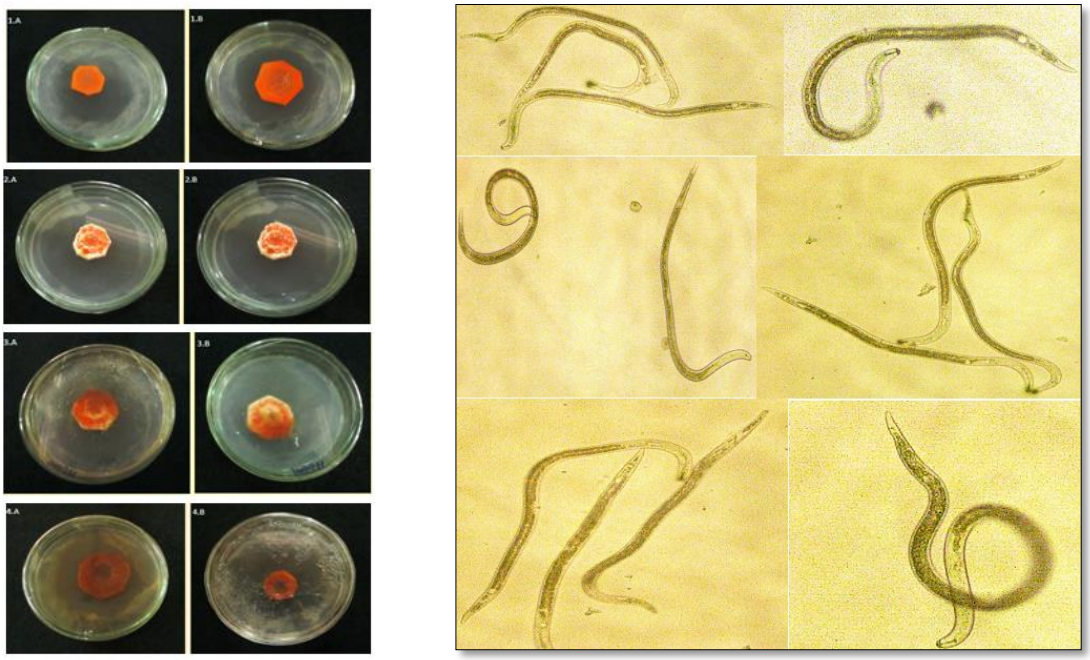


Figure 2: [A]*R. similis* carrot culture plates at various stages: (A1-A, B) shows a prepared and sealed carrot plates, (A2-A, B) shows the carrot plate after two weeks, when the callus is formed and plates are ready for inoculation with *R. similis*, (A3-A, B) shows the inoculated plates after two months where the infection is visible on the carrot plates, (A4-A, B) shows the plates after an incubation of more than two and half months. [B] *R. similis* nematode population as viewed under a compound microscope

1.2. DNA/ RNA extraction, amplification and sequencing of the selected genes for molecular characterization

Genomic DNA of *R. similis* was isolated based on Bolla *et al.* (1998) and total RNA using Total RNA Norgen Purification Kit (NorgenBiotek, Canada) according to manufacturer's protocol. First-strand cDNA was synthesized with ThermoScientificRevertAid Reverse Transcriptase in the presence of an oligo (dT) primer. Species confirmation from DNA and cDNA has been done using *R. similis* species specific primers for 18S (DNA) and actin (cDNA). The chromosomal non-coding genes, ITS, D3-Domain, 28S rRNA, 18SrRNA were amplified using the standardized PCR parameters for genomic DNA while mitochondrial genes - COX1 and ND1- and the control gene actin were amplified using the standardized PCR parameters for cDNA using the custom designed primers Table 2. The primers for actin was obtained from Haegeman *et al.*(2008).

Table 2: The details of primers designed for the study and their properties

Genes	Primer name	Primer 5'<-----Sequence----->3'	Length	Tm	GC
18S (546)	18S_rRNA_1_F	AACGAAAGTCTTCCGGTTCC	20	60.5°C	50.00%
	18S_rRNA_1_R	TACAAAGGGCAGGGACGTAA	20	60.5°C	50.00%
28 S (745)	28S_rRNA_2_F	GTACCGTGAGGGAAAGTTGA	20	57.6 °C	50.00%

	28S_rRNA_2_R	CGATTAGTCTTTCGCCCCTA	20	59.3°C	50.00%
COX1(999)	COX1_F	TTTTGGTTTTTACCTGCTGCT	21	58.9°C	38.10%
	COX1_R	TCCTTGTAAACCAGCAAAATG	21	57.8 °C	38.10%
ND1(701)	ND1_F	CTTTTAGCTTTAAGTCAGAATCGTC	25	57.3 °C	36.00%
	ND1_R	CGATAACGAGGTAAAACACTACGA	24	59.6 °C	41.70%
D3_DOMAIN (217)	D3_DOMAIN_2_F	AGTCATTGGGCGTTGAAAAC	20	60.0 °C	45.00%
	D3_DOMAIN_2_R	TTCGGACTTCCACCAGAGTT	20	59.7°C	50.00%
ITS (760)	ITS_F	GACTGAGCCGATTCGAGAAA	20	60.5°C	50.00%
	ITS_R	TCAGGTCGAATGATTTGGAA	20	59.0°C	40.00%

Amplified fragments were purified from agarose gels using ThermoScientificGeneJET Gel Extraction Kit (Thermo Fisher Scientific Inc., MA USA). Eluted PCR products were then ligated in the pTZcloning vector (ThermoScientific) using standard protocol. Freshly prepared competent cells of *Escherichia coli* JMI-109 were transformed with recombinant plasmids and positive clones were selected and grown in LB medium. Plasmids were isolated using ThermoScientificGeneJET Plasmid Extraction Kit (Thermo Fisher Scientific Inc., MA USA) and sequenced at M/s Scigenom, Kochi, India.

1.3. Sequence data analysis

Both the forward and reverse sequences obtained were checked for vector contamination using VecScreen (https://www.ncbi.nlm.nih.gov/tools/vecscreen/). Any contamination found was removed by trimming using BioEdit tool (Hall, 1999). The trimmed forward and reverse sequences were then aligned using PRABI-Doua: CAP3 Sequence Assembly Program (Huang and Madan, 1999). Putative protein sequences for COX1 and ND1 were obtained by translating the gene sequences using the EMBOSS program TRANSEQ (Rice *et al.,*2000). Conserved domains were predicted by NCBI Conserved Domain Database and InterProScan (Quevillon *et al.,* 2005). Protein structure analysis was done with Expasy-Protparam tool (Gasteiger *et al.,*2005) to analyze the various physical and chemical parameters of protein sequences. The computed parameters were theoretical isoelectric point (PI), extinction coefficient, molecular weight, amino acid composition, atomic composition, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).The molecular sequences were compared with corresponding proteins of other nematode species available in the GenBank sequence database using the basic local alignment search tool (BLASTN),a homology search program(Altschul *et al.,* 1990). Sequences were also aligned using the ClustalW implementation in Geneious version 9.1.3. The sequence diversity analysis was performed and G/C content, number of polymorphic sites, average pairwise nucleotide, and difference per site (p), mutations, insertion deletion events (InDel), sequence divergence and sequence conservation along with the conservation threshold were identified using DNAsp.5 (Librado and Rozas, 2009) and MEGA6 (Tamura *et al.,* 2007). The average non-synonymous/synonymous rate ratio (dN/dS) was calculated with KaKs Calculator Version 2.028 (Wang *et al.,* 2010) to infer direction and magnitude of natural selection.

1.4. Phylogenetic analysis using MLST

For phylogenetic analysis, sequences of all the available sequences of the six genes were downloaded from NCBI-GenBank database for all the PPNs (Supplementary File S1). For COX1 and ND1 genes, their protein sequences were used. Bayesian analysis was performed to confirm the tree topology for each gene separately using MrBayes-3.2.6 (Ronquist *et al.,* 2011). Two simultaneous runs consisting of 500,000 generations, each of four chains, were performed in all analyses. Trees were sampled after every 500 generations. Stationarity was considered to have reached when the average standard deviation of split frequencies was less than 0.002. Markov Chain Monte Carlo (MCMC) methods were used within a Bayesian framework to estimate the posterior probabilities (pp) of the phylogenetic trees using the 50% majority-rule.

For generating MLST, the aligned genes were concatenated using Geneious version 9.1.3.ND1 gene sequences were not considered for the generation of MLST, as sequences of this gene were not available for other nematodesin the GenBank database. The concatenated genes from all the organisms were further aligned using ClustalW implementation within Geneious. The best fitting evolutionary model of each partition using Akaike information criteria (AIC) in the individual DNA sequence data as well as the concatenated gene sequence was determined by the software jModeltest version 2.1.7 (Darriba*et al.,* 2012). The Akaike-supported model, the proportion of invariable sites and the gamma distribution shape parameters and substitution rates were used in phylogenetic analyses. A GTR model with gamma-distributed substitution rates was selected for all datasets along with an estimated proportion of invariant sites. A Yule prior was used and the remaining model parameters were estimated from the data. A Bayesian search was performed on each dataset for 1000000 generations logging every 1000 using BEAST v1.6.2 (Drummond *et al.,*2012) for the concatenated genes. Phylogenetic relationship was further inferred using various methods like maximum parsimony using the dnapars package inPAUP version 4.0a150 (Swofford, 2002), maximum likelihood usingPhylip 3.695 (Felsenstein, 2014), Bayesian method using MrBayes-3.2.6 (Ronquist *et al.,* 2003) and neighbor-joining using MEGA5 (Tamura *et al.,* 2007). Finally, the trees obtained from all the four methods were used to draw the consensus tree. The majority-rule consensus of the bootstrap replicate trees was calculated using consense and seqboot package in Phylip 3.695 (Felsenstein, 2014).

1.5. Intra-species variation analysis

For inferring the intraspecies variation between the Indian population of *R. similis*and *R. similis*from various geographical locations, sequences of all the six genes that were available for *R. similis*within NCBI-GenBank database were downloaded. Only the sequences for which geographical location of isolation was provided were considered for the analysis. Since the two mitochondrial genes COX1 and ND1 did not have enough sequences available for comparison, these sequences were not considered for further analysis. In the genes 18S rRNA, 28S rRNA and D3 expansion domain sequences were present in NCBI-GenBank database but most of them were from the same locations. Therefore, due to lack of variation in location of these genes, they were also not considered for the study. A total of 82 sequences *R. similis* ITS regions from known locations were downloaded. ITS sequences for two additional species of *Radopholus* namely *R. arabocoffeae* (6 sequences) and *R. duriophilus* (4 sequences) were included to

provide a root to networks and estimates of phylogeny. Sequences were aligned and trimmed using the ClustalW implementation in BioEdit tool (Hall, 1999). The final alignment used for network analysis comprised of 93 sequences and was 613 base pairs in length.

Haplotype distribution study was done using DnaSP v5 to divide the 93 sequences into different haplotypes. The relationship between population and their geographical distribution based on these haplotypes were estimated using TCS networks (Clement *et al.,* 2002) method implied in the POPART v1.7 program (Leigh and Bryant, 2015). For the ITS gene sequences Bayesian analysis was also performed using MrBayes-3.2.6 tool to calculate the phylogenetic relation with other downloaded sequences. Two simultaneous runs, each of four chains, were performed in all analyses. Each run consisted of 200,000 generations, and trees were sampled every 200 generations. Stationarity was considered to be reached when the average standard deviation of split frequencies was less than 0.002. Markov Chain Monte Carlo (MCMC) methods were used within a Bayesian framework to estimate the posterior probabilities of the phylogenetic trees using the 50% majority-rule.

2. TRANSCRPITOME SEQUENCING, ASSEMBLY AND ANNOTATION OF *R. SIMILIS*

2.1. RNA isolation and sequencing

For RNA isolation nematodes were cultured and isolated as stated in 1.1. and around 10,000 nematodes were collected. RNA was isolated using Total RNA Norgen Purification Kit(NorgenBiotek, Canada) according to manufacturer's protocol. Library preparation and high-throughput sequencing services were performed at the Bionivid Technology Pvt Ltd., Bengaluru, India.

2.2. *De novo* transcriptome assembly, annotation and quantification

Prior to assembly, raw reads from these libraries were filtered using Trimmomatic (https://github.com/usadellab/Trimmomatic) to remove low quality reads. Picard (https://github.com/broadinstitute/picard/releases/tag/2.25.5) was used to find and remove the PCR duplicates present in the raw data. The remaining reads were paired and orphan reads were discarded. High quality paired-end reads as well as raw reads were used as input for transcriptome assembly. *De novo* transcriptome assembly was completed using the de Bruijn graph-based tool Trinity (https://github.com/trinityRNA-seq/trinityRNA-seq) on raw data, trimmed data and PCR duplicate removed data. As part of the assembly process, an *in silico* read normalization step was performed. Assembly quality was then assessed by using *in silico* tool BUSCO v3 (https://gitlab.com/ezlab/busco) at default parameters. The transcriptome was annotated following the established Trinotate pipeline (https://github.com/Trinotate/Trinotate.github.io). Homology searches were performed against the protein sequences contained in Genbank and UniProt databases using BLASTX at an 'e' value cutoff of 1e-5. Transcripts were translated into protein using TransDecoder, a component of Trinity. HMMER (http://hmmer.org/) and Pfam (http://pfam.xfam.org/) databases were used to predict protein domains contained within each transcript. After conceptual translation from the predicted coding domains, the functions of the potential proteins were predicted using InterProScan (https://www.ebi.ac.uk/interpro/download/), using the default parameters. According to their

homology with conserved domains and with protein families in NCBI nr database, protein sequences were assigned to three gene ontology (GO) categories, including molecular function, cellular component and biological process using Blast2Go (https://www.biobam.com/download-blast2go/) (Götz *et al.,* 2008). The pathway analysis of inferred proteins was carried out using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (https://www.genome.jp/kegg/pathway.html) (Kanehisa and Goto, 2000). SSR prediction was done using perl-based script MISA (Beier *et al.,* 2017).

2.3. Prediction of Excretory/Secretory (ES) proteins

ES proteins were identified using integration of six bioinformatics tools; SignalP (Armenteros*et al.,* 2019), Phobius (Käll*et al.,* 2017), TMHMM (Möller*et al.,* 2001) and TargetP (Emanuelsson*et al.,* 2000), PredGPI (Pierleoni*et al.,* 2000) and Protcomp (Protcomp, 2000). Several variations ofthese computational approach for the prediction of classically secreted ES proteins have been successfully applied earlier to different parasitic organisms. SignalP v5 and Phobius v1.01 were used for predicting classically secreted proteins. For SignalP prediction,the parameter was set for eukaryote organisms, truncation of protein sequence at 70 residues and D-cut-off values as default. Phobius was used with default options. Proteins predicted as secretory from SignalP and Phobius were passed to TMHMM with default parameters to check for proteins containing transmembrane domains. TargetP v1.1 was used for identifying mitochondrial targeting proteins at 95% of specificity. Mitochondrial proteins predicted by TargetP were removed from the protein dataset and remaining proteins were then passed to PredGPI v1.1 for detecting GPI-anchored proteins. The GPI-anchored proteins were eliminated from protein dataset and remaining proteins were checked for other localized proteins using Protcomp. Proteins containing localization signals were then removed and finally, the remaining proteins were considered as ES proteins for further analysis.

2.4. Functional and comparative annotation of ES proteins

The predicted ES proteins were functionally annotated using various bioinformatics tools. The ES proteins, coded for carbohydrate active enzymes (CAZymes) were identified using dbCAN web server (version 5) (Yin *et al.,* 2012), with an E-value threshold of 1e-05. The ES proteins were independently assessed for orthologs against the Cluster of Orthologous Groups of proteins (COG) database (Tatusov*et al.,* 2000) using BLAST with permissive (E-value: 1e-10) search strategies. Finally, the ES proteins were predicted to have pathogenic function by BLAST analysis of the Pathogen Host Interaction (PHI) database (Winnenburg*et al.,* 2006) (identity >25, E-value: 1e-10).We compared the ES protein sets of *R. similis* against those of 32 other species using Blastp with default parameters to detect putative homologs based on best blast hit approach. Species whose whole genomes have been annotated to a quality level allowing a reliable prediction of protein-coding genes were selected for this study. Our selection of species comprised of total 18 species that were listed as blacklisted/non-target species, involved three free living nematodes, one animal parasitic nematode, nine plant parasitic nematodes, four insects, eight vertebrates (including mammals, ray-finned fishes, amphibian and sauropsida), four fungi and three plant species. Whenever an ES protein had a predicted homology in these blacklisted/non-target species, the protein was discarded from the downstream analysis. The ES

proteins non-homologous to the blacklisted/non-target proteome were further screened for RNAi phenotypes in *C.elegans*.

3. IDENTIFICATION AND SEQUENCING OF NOVEL TARGET GENES

3.1. PCR amplification, cloning and sequencing of genes

Nematode isolation, RNA isolation and cDNA library preparation were done as mentioned in 1.1 and 1.2. Nine target genes selected from the predicted ES proteins were amplified and cloned. The selected genes were abhydrolase domain-containing protein (Rs-AB), immunodominant hypodermal antigen Ac16 (Rs-Ac16), astacin-like metalloendopeptidase (Rs-AsLM), carbonate dehydratase, eukaryotic-type (Rs-CDh), EGF-like domain protein (Rs-ElDP), vacuolar protein sorting-associated protein 45 (Rs-VpSA), glycoside hydrolase, superfamily domain-containing protein (Rs-GH), pectate lyase (Rs-PL) andreticulocalbin-2 (Rs-ReCa2).The primers for the above genes were custom designed from assembled *R. similis* transcriptome data using primer designing tools whereas primer for actin was obtained from the published literature (Haegeman *et al.,* 2008). Primer sequences used for amplification of each gene are listed in Table 3.

Table 3: Primers used for amplification of target genes in *Radopholus similis*

Gene	Contig id	Primer name	Primer sequence	Length	Tm	Gc
Rs-AB	TRINITY_DN3434_c0_g1_i1	ABHYDROLASE_F	TGACTGCGGCAAACCTAGAC	1048	60.32	55
		ABHYDROLASE_R	CTCTTTGCGCACTCTCCTCT		59.75	55
Rs-Ac16	TRINITY_DN27926_c0_g1_i1	AC16_F	CTATGGCGCCTTCTCTGCTT	620	60.18	55
		AC16_R	TCGTTATAGTCCCAAAATGCCA		57.78	40.91
Rs-AsLM	TRINITY_DN3080_c0_g1_i1	ASTACIN_F	GTTTGGATTTGGCCGAGCAG	984	60.11	55
		ASTACIN_R	GTCGACCTTGCTACTGTGCT		60.04	55
Rs-CDh	TRINITY_DN20759_c0_g1_i24	CARBONIC_F	TAAACACAGCCTTCGTCCCC	1310	59.96	55
		CARBONIC_R	GGGATTGCGCACGTCATATT		59.06	50
Rs-ElDP	TRINITY_DN13188_c0_g1_i1	EGF_LIKE_F	GACCGTTTCCCCGACTTCTT	1028	59.97	55
		EGF_LIKE_R	CATGCCATCCAAACCACGAC		59.83	55
Rs-VpSA	TRINITY_DN13185_c0_g1_i1	EXPAN_B2_F	GTAGGTGATGGTTGCTCCCC	1106	60.11	60
		EXPAN_B2_R	GCGCACTCTGTCCAAACAAG		60.04	55
Rs-GH	TRINITY_DN22159_c3_g1_i5	GLY_HYDRO_F	GCTTCAGCGACAGTGGTAGA	957	59.2	55
		GLY_HYDRO_R	GGGAGATGTCGCCACTTTTA		60.1	50
Rs-PL	TRINITY_DN11189_c0_g1_i2	PEC_LYS2_F	ATTCGGGGACTCTTGTGCAAC	932	60.88	52.38
		PEC_LYS2_R	ATGAGCCTGCACTTACTGCTT		60	47.62
Rs-ReCa2	TRINITY_DN19527_c0_g1_i3	RETICULO_F	AGCGTCAATTGTCCGTCCTT	1285	59.97	50
		RETICULO_R	GCAGCCGGCACAGAAATATG		59.97	55

Amplified fragments were purified from agarose gels using ThermoScientificGeneJET Gel Extraction Kit (Thermo Fisher Scientific Inc., MA USA). Eluted PCR products were then ligated in the pTZcloning vector (ThermoScientific) using standard protocol. Freshly prepared competent cells of *Escherichia coli* DH5α were transformed with recombinant plasmids and positive clones were selected based on blue white colony discrimination and grown in LB medium. Plasmids were isolated using ThermoScientificGeneJET

Plasmid Extraction Kit (Thermo Fisher Scientific Inc., MA USA) and sequencing was done at M/s Scigenom, Kochi, India.

3.2. Sequence data analysis

Both the forward and reverse sequences obtained were checked for quality and pre-processed as describe in 1.3 and submitted to NCBI.Each of the target genes obtained by sequencing were translated into putative protein sequences using the EMBOSS program TRANSEQ (Rice *et al.,*2000) for further *in silico* analysis. Protein structural analysis with the target protein sequence was done with Expasy-Protparam tool (Gasteiger *et al.,* 2005) for analyzing various physical and chemical parameters of target proteins.The computed parameters were theoretical isoelectric point (PI), molecular weight, extinction coefficient, amino acid composition, estimated half-life, atomic composition, grand average of hydropathicity (GRAVY), instability index and aliphatic index.

3.3. Motif characterization and evolutionary analysis

Conserved domains were predicted by searching the matches against protein signature databases such as NCBI Conserved Domain Database (CDD) (Marchler-Bauer and Bryant, 2004) and InterProScan (Quevillion *et al.,* 2005). The multiple sequence alignment of target proteins with corresponding known parasitic nematode genes were created with Clustal omega program (Madeira et al.,2019). Conserved motifs and their relative positions of target proteins were predicted using MEME Version 5.1.0 (Bailey*et al.,* 2009). Motif of a sequence is a pattern of amino acid that occurs repeatedly in a one or more sequences of a group or a protein family (Bailey *et al.,* 2015).

The evolutionary history was inferred from the aligned protein sequences of each gene using Bayesian algorithm, maximum parsimony and maximum likelihood analysis. Bayesian analysis was performed to confirm the tree topology for each gene separately using MrBayes-3.2.6 (Ronquist *et al.,* 2011). Two simultaneous runs, each of four chains, were performed in all analyses. Each run consisted of 500,000 generations and trees were sampled every 500 generations. Stationarity was considered to be reached when the average standard deviation of split frequencies was less than 0.002. Markov Chain Monte Carlo (MCMC) methods were used within a Bayesian framework to estimate the posterior probabilities (pp) of the phylogenetic trees using the 50% majority-rule. Phylogenetic relationship was further inferred using various methods like maximum parsimony using the dnapars package Phylip 3.695 (Felsenstein, 1993), maximum likelihood using PAUP version 4.0a150 (Wilgenbusch and Swofford, 2003). Finally, the trees obtained from all the three methods were used to draw the consensus tree for each target protein. The majority-rule consensus of the bootstrap replicate trees was calculated using consensus and seqboot package in Phylip 3.695 (Felsenstein, 1993).

4. EXPERIMENTAL VALIDATION OF NOVEL TARGET GENES

4.1. Synthesis of target dsRNA and soaking studies

Target dsRNAs were generated from amplicons corresponding to coding regions of functional protein domains of the above genes and are designated as "dsRNAs" throughout the manuscript. The sizes of the target genes used to generate dsRNAs ranged from 232 to 500 bp (Supp.Table 1). The amplicons were obtained from cDNA generated from total RNA of mixed stages of *R. similis* as described by Iqbal et al. (2016). The PCR products were purified using Purelink PCR purification kit following manufacturer's protocol. Target dsRNAs were synthesized using the HiScribe T7 In vitro Transcription Kit (New England Biolabs, Australia) as described by Iqbal et al. (2016) from templates prepared from the purified PCR product using custom designed primers (Table 4). DNAseI digestions of the dsRNA obtained were performed followed by Na-acetate and ethanol purification.

Table 4: T7 primers used for amplification of dsRNA genes in *Radopholus similis*R01

Gene	Primer name	Primer sequence	Length	Start	Tm	Gc
Rs-AB	Rs-AB_1F	taatacgactcactatagggCACCGAGTTCGGACAGTTCT	500	13	60.298	55
	Rs-AB_1R	taatacgactcactatagggCGAAACATTTCACCCTGCTT		512	60.11	45
Rs-Ac16	Rs-Ac16_1F	taatacgactcactatagggCTGAAATTCGCCATTGCTG	232	7	60.35	47.368
	Rs-Ac16_1R	taatacgactcactatagggCCACAATTTGTTAGCCAGCTC		238	59.761	47.619
Rs-AsLM	Rs-AsLM_1F	taatacgactcactatagggACGCGTAGAACTAGTCCCGA	497	105	59.898	55
	Rs-AsLM_1R	taatacgactcactatagggGCTGGCTGACCTTTTCAAAC		601	59.859	50
Rs-CDh	Rs-CDh_1F	taatacgactcactatagggATTCCCGTTCAGCTGGAGTT	253	22	61.019	50
	Rs-CDh_1R	taatacgactcactatagggGGTAGTGCAGGGCATTCATT		274	59.962	50
Rs-ElDP	Rs-ElDP_1F	taatacgactcactatagggATTGCGTTCCTCCATAGCC	386	24	60.051	52.632
	Rs-ElDP_1R	taatacgactcactatagggTGATCATGGACAGAAGCGAG		409	59.82	50
Rs-VpSA	Rs-VpSA_1F	taatacgactcactatagggTGATCATGGACAGAAGCGAG	464	254	59.942	50
	Rs-VpSA_1R	taatacgactcactatagggTTTGAGGGACTCCAAACAGG		717	60.081	50
Rs-GH	Rs-GH_1F	taatacgactcactatagggAAGCTGGCATCAGCAACTTT	496	215	60.022	45
	Rs-GH_1R	taatacgactcactatagggTTCTGTCTCATCTGGACCCC		710	60.048	55
Rs-PL	Rs-PL_1F	taatacgactcactatagggGCCGACGTCTTGTAGTTGGT	498	83	60.179	55
	Rs-PL_1R	taatacgactcactatagggCGTTTCCAACGTGATCATTG		580	59.964	45
Rs-ReCa2	Rs-ReCa2_1F	taatacgactcactatagggCCCACGTGATAATACCGTCC	499	278	60.074	55
	Rs-ReCa2_1R	taatacgactcactatagggCATCACATCACTTCGTTGGC		776	60.12	50

4.2. Assessment of Target Gene after soaking

Study by Wang et al. (2016) reported that neuro-stimulant is not required to induce *R. similis* to ingest dsRNA in a soaking solution via many of its organs in both vermiform nematodes and egg shells hence

no neurostimulants were used for dsRNA uptake. Duration of the silencing effect was studied by Rosso et al. (2005) and have reported the highest silencing effect was obtained 20 h after soaking. This effect was reduced 44 h after soaking and was no longer detectable 68 h after soaking. Therefore, for each silence assay, about 500 mixed stage nematodes were washed with SDwater, soaked in dsRNA solution (2.0 mg/mL), and shaken slightly in a dark rotary incubator at 25°C for 4 h, 24 h, 44 h, respectively. Meanwhile, untreated nematodes were used as a blank control. To demonstrate uptake of the soaking solution, 100 nematodes were soaked with 1 mg mL^{-1} Fluorescein Isothiocyanate (FITC) and 1 mg mL^{-1} dsRNA; the presence of the FITC in the stylet, gut, and the body of the nematode was used as a sign of uptake of the solution, more likely through ingestion. The activity and behavior of the nematodes from the three biological replicates for each dsRNA and the no-dsRNA controls were observed with a Leica DM 5000B research microscope under bright-field, whereas those soaked with FITC were assessed using the FITC filter (450–480 nm) on the microscope.

Approximately 1000 mixed-stage nematodes were washed with DEPC water and then soaked in dsRNA solution (2.0 mg/ml) with gentle agitation in a dark rotary incubator (100 rpm, 25 °C) for 4, 24 and 48 h, respectively. Total RNA was extracted from 100 nematodes in each treatment group after they were washed with DEPC water, representing each of the three biological replicates of the treatments. cDNA synthesized from RNA extracted was utilized for detect the silencing efficiency through quantitative PCR. Replicates of qPCRs were done with a 1:10 dilution of the cDNA in 20 µL using 10 µM each of primer pairs with the QuantiFast SYBR Green PCR Kit (Qiagen). The cycling conditions were an initial denaturation for 10 min at 95◦C followed by 40 cycles of 95◦C for 15 s, 55◦C for 20 s, and 72◦C for 30 s in a Qiagen Rotor Gene Q. Primers used were custom designed from amplified and sequenced target genes and were selected based on their specificities as determined by melting curve (Table 5).

Table 5: Primers used for assessment of target gene knockdown

Gene	Primer name	Primer sequence	Length	Start	Tm	Gc
Rs-AB	Rs-AB_qrt_F	GTTTCGGCTGCTCTTCGTAT	129	687	59.5	50.0
	Rs-AB_qrt_R	TAGAGGGACGGACTGAATGG		815	60.1	55.0
Rs-Ac16	Rs-Ac16_ qrt_F	CCTGCATCTTTTCCTTTTCG	113	50	59.8	45.0
	Rs-Ac16_ qrt_R	CAGTCAGAGGATGTGCAGGT		162	58.8	55.0
Rs-CDh	Rs-CDh_ qrt_F	AACGAGCTGCACTTTGTGAA	149	52	59.6	45.0
	Rs-CDh_ qrt_R	AGTGCGTACTTTCCGCTCAG		200	60.6	55.0
Rs-VpSA	Rs-VpSA_ qrt_F	AAAGACTGGCGGAGGAATTT	142	173	60.1	45.0
	Rs-VpSA_ qrt_R	GCCTCGTATGTCCACTGGTT		314	60.0	55.0
Rs-GH	Rs-GH_ qrt_F	GCCGACTACAGCTGTCTGAA	135	160	59.2	55.0
	Rs-GH_ qrt_R	AAAGTTGCTGATGCCAGCTT		294	60.0	45.0
Rs-PL	Rs-PL_ qrt_F	GAGGCCACGTAGTTGGTGTT	103	110	60.0	55.0
	Rs-PL_ qrt_R	ACTACGGCGACTCTTTGAGC		212	59.6	55.0

Rs-ReCa2	Rs-ReCa2_ qrt_F	CCGTCTTTGTTTGCGTCTTT	118	34	60.3	45.0
	Rs-ReCa2_ qrt_R	GCGACATGAAGTTGAGCAAA		151	60.0	45.0

Fold changes in target gene expression were calculated using the comparative Ct method (ΔΔCt); the values obtained represented a mean of the three biological replicates of the soaking experiments (Livak and Schmittgen, 2001). A total of 30 female nematodes from each replicate were inoculated onto carrot disks and cultured for 50 d at 25°C, and then the reproductive rates (reproductive rate = final nematodes ⁄ initial nematodes) of the nematodes were calculated.

5. DEVELOPMENT OF BIOINFORMATICS PIPELINE

RNA-seq data analysis is a multi-steps process, mostly all of them being computationally intensive.Although, several tools dealing with individual steps are available, integrated pipelines covering all important aspects of RNA-seq analysis is rarely seen. RNA-seqAnalysis isan automated R/Bioconductor-based pipeline that offers an end-to-end analysis for RNA-seq data.

- The workflow covers following steps of RNA-seq analysis: quality check, filtering and trimming, assembly, differential gene expression, annotation, splice variant identification (Figure 3).
- The workflow successfully integrates different R and Bioconductor packages for different steps of RNA-seq analysis (Table 6).
- This pipeline was tested with *Caenorhabditis elegans* SRP099636 (Ma *et al.,* 2016) dataset which have been randomly sampled to 90,000 PE reads.
- The working has been demonstrated and explained in form of vignette (user manual).
- The pipeline is uploaded in Git-hub as a bioconductor package for review and is freely available for download.

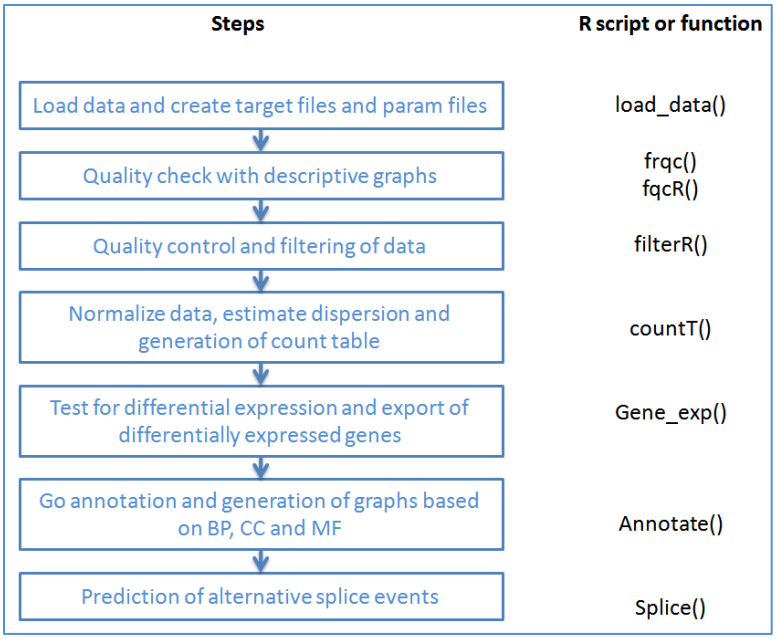

Steps
R script or function
Load data and create target files and param files
load_data()
Quality check with descriptive graphs
frqc()
fqcR()
Quality control and filtering of data
filterR()
Normalize data, estimate dispersion and generation of count table
countT()
Test for differential expression and export of differentially expressed genes
Gene_exp()
Go annotation and generation of graphs based on BP, CC and MF
Annotate()
Prediction of alternative splice events
Splice()

Figure 3. R functions associated with each step when running the pipeline

Table 6: Bioconductor packages used for each function

S. No.	Function	Bioconductor packages used	Reference
1	frqc	Rqc	Souza and Carvalho 2018
2	fqcR	qrqc	Buffalo 2012
3	filterR	Short read	Morgan et al.,., 2009
4	countT	systemPipeR, easyRNA-seq, GenomicAlignment, GenomicFeatures, ggplot2, AnnotationDbi,	Backman and Girke 2016; Delhomme et al.,.,2012; Lawrence et al.,., 2013;Hadley, 2016;Pages et al.,., 2013
5	Gene_exp	edger, DESeq2, systemPipeR, easyRNA-seq, gplot, ggplot2, pheatmap, RcolorBrewer, lattice, SummarizedExperiment	Robinson et al.,., 2010; Love et al.,., 2014; Backman and Girke 2016; Delhomme et al.,., 2012;Warnes et al.,., 2016; Hadley 2016; Kolde 2012;Neuwirth and Neuwirth 2011; Sarkar et al.,., 2015; Morgan et al.,., 2017
6	Annotate	biomaRt, systemPipeR	Durinck et al.,., 2005, Backman and Girke 2016
7	Splice	Aspli, GenomicFeatures, AnnotationDbi	Schlaen et al.,., 2015, Lawrence et al.,., 2013, Pagès et al.,., 2018

RESULTS AND DISCUSSION

1. MOLECULAR CHARACTERIZATION OF *R. SIMILIS*

1.1. DNA/RNA extraction, amplification and cloning

Genomic DNA isolation from *R. similis* resulted in a yield of 200ng/µl and the yield of total RNA was 64.9 ng/µl. First-strand cDNA was synthesized and species confirmations were done by amplifying cDNA with *R. similis* specific actin primer. The ribosomal non-coding genes, ITS, D3-Domain, 28S rRNA, 18S rRNA genes and mitochondrial genes - COX1 and ND1and the control gene actin on assembly resulted in sequences ranging from 1380-1560 nucleotides, which got reduced after vector removal (Table 7). The obtained partial sequences were submitted in NCBI with accession numbers KU530150- KU530152, KX396030, KX378136 and KX420702.

Table 7: Nuclear and mitochondrial genes amplified and sequenced from an Indian isolate of *R. similis*R01

Gene	Seq length after assembly	Seq length after vector removal	GenBank id
COX1	1445	899	KX378136
ND1	1639	703	KX420702
ITS	1560	762	KX396030
18S	1455	523	KU530150
28S	1380	742	KU530151
D3	1535	219	KU530152

1.2. Sequence data analysis

The sequences obtained from *R. similis* R01 population were >98 identical to that of *R. similis* sequences present in GenBank when performed blast search, showing highly reliable sequencing results. In case of COX1 and ND1, the sequences were blasted against the available *R. similis* mitochondrial genome. The highest percentage identity was obtained for 18S rRNA genes. Details of blast search for each gene are provided in Table 8.

Table 8: Results of Blast search using nuclear and mitochondrial genes of Indian isolate of *R. similis* R01

Genes	Blast hit against	Identity percentage	Query coverage	Gap	Mismatch	E value
COX1	FN313571.1	99.00%	99.00%	0	1	0
ND1	FN313571.1	98.00%	100.00%	0	5	0
ITS	KJ845637.1	99.00%	99.00%	0	1	0
18S	KJ845641.1	100.00%	99.00%	0	0	0
28S	KF234211.1	99.00%	99.00%	0	4	0
D3	JQ782261.1	99.00%	99.00%	0	1	$1e^{-106}$

Prediction of putative protein sequences for COX1 and ND1 by translating the cDNA sequences using the EMBOSS program TRANSEQ, resulted in a polypeptide of 196 and 126 amino acid residues, respectively. The amino acid composition of COX1protein showed the highest number of leucine and serine residues (24) whereas phenyl alanine was completely absent. The amino acid composition of putative protein of ND1 showed the presence of highest number of leucine residues (29) while phenyl alanine residues were completely absent. The instability index of proteins was found to be within the range of stability as

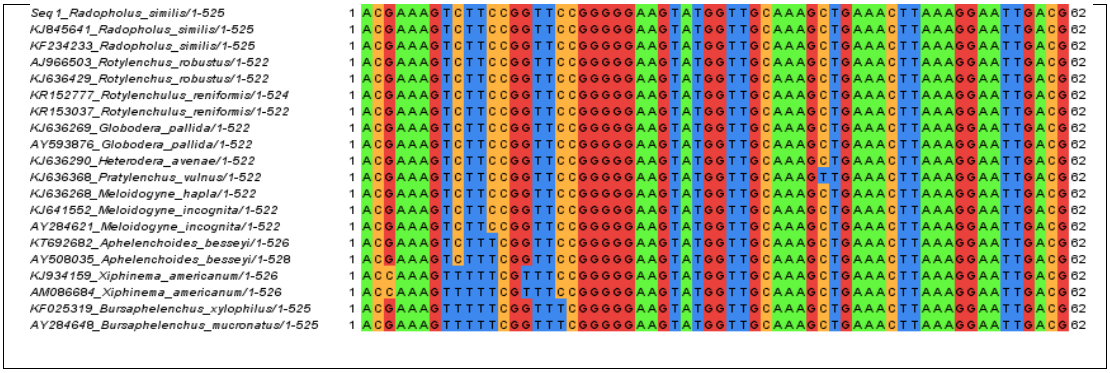
all values were found to be less than 40, and any protein whose instability index is smaller than 40 is predicted as stable. The half-life is a prediction of the time it takes for half of the amount of protein in a cell to disappear after its synthesis in the cell. For both COX1 and ND1 the half-life was predicted to be >20 hours (yeast, in vivo).The results obtained from Expasy-Protparam tool is shown in Table 9. Analysis by NCBI Conserved Domain Database for the protein Cox1 revealed the presence of conserved multi domain Cytochrome C and Quinol oxidase polypeptide I and similarity with the superfamily Heme-copper oxidase subunit I. The putative protein sequence of ND1 gene showed the presence of NADHdh super family.

Table 9: Results from Expasy-Protparam tool showing various physical and chemical parameters of protein sequences COX1 and ND1 from Indian isolate of *R. similis* R01

Target	No. of amino acids	Molecular Weight	Isoelectric Point (pI)	Instability Index	Aliphatic Index	GRAVY	Negative Residues	Positive Residues
COX1	192	21041.16	6.55	24.47	103.02	0.919	8	7
ND1	126	15164.52	9.41	25.11	133.81	1.172	5	10

1.3. Nucleotide polymorphisms and amino acid changes

Multiple sequence alignment (MSA) was done using the new sequences of *R. similis* R01 against set of sequencesfrom different PPNs downloaded from NCBI-GenBank. The nucleotide sequence of ITS, D3 expansion domain, 18S rRNA and 28S rRNA were taken for MSA. Since the protein sequence of COX1 and ND1 were available, protein sequences were considered. The percentage identity score for all the genes showed highest score against other *R. similis* sequences, but the lowest score obtained were with different organisms, mostly *Xiphinema americanum* (Supplementary Table S3). The alignment resulted in the highly conserved regions for the 18S rRNA gene and was the least conserved regions were present in the ITS1-5.8S-ITS2 alignment (Figure 4).



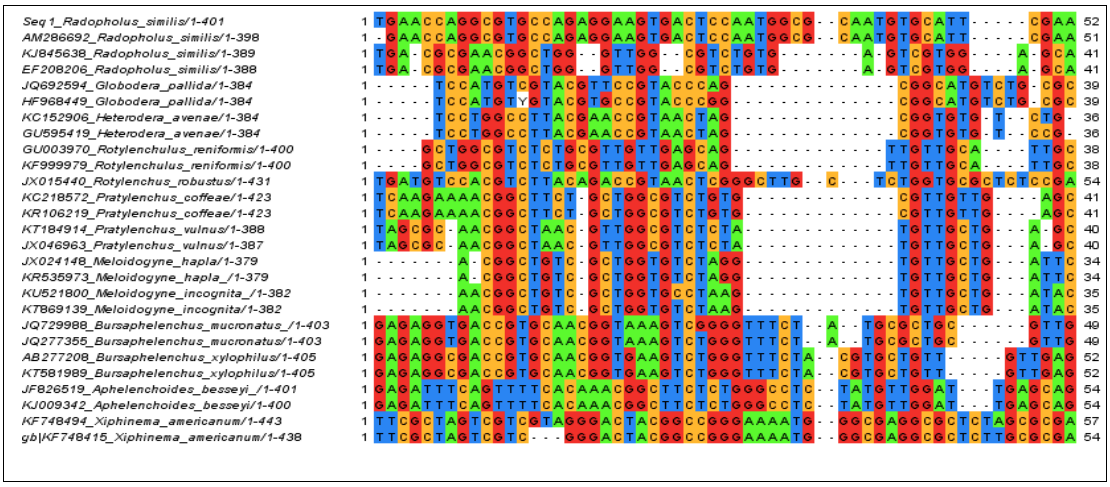


Figure 4:Multiple sequence alignment of (Top) 20 sequences of 18S rRNA gene and (Bottom) 27 sequences of ITS1-5.8S-ITS2 regions from different plant parasitic nematodes

The polymorphism analysis yielded information on G/C content, number of polymorphic sites, average pairwise nucleotide, and difference per site (p), mutations, insertion deletion events (InDel), sequence divergence and sequence conservation of all the six genes (Table 7). It showed that the sequence conservation was lowest in the case of ITS1-5.8S-ITS2 and highest in case of 18S rRNA genes. GC content for all the ribosomal genes were around 50% showing high stability of the gene, whereas in the case of mitochondrial gene the GC content was low, as reported by Jacob et al. (2009). Insertion/deletion was lowest in the case of COX1 and highest in case of ITS1-5.8S-ITS2. The highest percentage of polymorphic sites and divergence was in ITS1-5.8S-ITS2and the lowest in the case of 18S rRNA.

The overall transition/transversion bias (R) was calculated for all the genes and the highest value for R was found in D3 expansion domain, followed by 28S rRNA, 18S rRNA, COX1, ITS1-5.8S-ITS2 and the lowest value of R was estimated in ND1 (Table 10). R stands for the ratio of the number of transitions to the number of transversions for a pair of sequences and is equal to 0.5 when there is no bias towards either transitional or transversional substitution because, when the two kinds of substitution are equally probable, there are twice as many possible transversions as transitions.

Table 10: Maximum composite likelihood estimates of the pattern of nucleotide substitution in nuclear and mitochondrial genes of Indian isolate of *Radopholus similis* R01

Genes	R	Transition/Transversion rate ratios		Nucleotide frequencies (%)			
		k1 (purines)	k2 (pyrimidines)	A	T/U	C	G
COX1	1.406	1.616	6.427	30.78	27.44	9.45	32.33
ND1	0.36	1.628	0.97	20.85	61.37	6.92	10.85
ITS	0.963	1.689	2.207	22.91	26.69	22.60	27.80
18S	1.152	0.98	3.993	24.72	25.90	28.26	21.13
28S	2.62	5.081	5.573	21.62	22.68	22.70	33.01
D3	3.396	2.251	14.786	27.49	19.94	21.39	31.18

Ka/Kswas calculated for both the protein coding genes, COX1 and ND1. A Ka/Ksvalue close to 1 indicates either neutral evolution at the protein level or an averaging of sites under positive and negative selective pressures and Ka/Ks value less than 1 implies purifying or stabilizing selection (acting against change). The Ka/Ks value obtained for COX1 using a set of 51 protein sequences across PPNs was 0.653734 and in case of ND1 analysis using 24 protein sequences across PPNs resulted in a Ka/Ks value 0.800177, indicating purifying or stabilizing selection on these genes.

1.4. Phylogenetic analysis

A final alignment of 2467 nucleotides of concatenated sequence of 18S rRNA, 28S rRNA, D3-expansion domain, ITS1-5.8S-ITS2 and COX1 was used in the analyses. The analysis involved 11 different species of plant-parasitic nematodes. Phylogenetic relationships within and between PPNs species were carried out on the concatenated sequences by means of Bayesian Inference (BI), Neighbor-Joining (NJ) with LogDet distances, Maximum Likelihood (ML) and Maximum Parsimony (MP) without using any outgroups. All the four trees obtained were used to generate the consensus tree. The phylogenetic trees as inferred from consensus-based phylogeny are given in Figure 5. The results showed that species formed a highly supported clade with *R. similis* sequences downloaded from NCBI in all the analysis. The *R. similis* clade showed a close relationship with other PPNs like *Rotylenchulus reniformis, Heterodera avenae, Globodera pallida, Pratylenchus vulnus, Meloidogyne incognita* and *Meloidogyne hapla* as previously reported by (Holterman *et al.,* 2009; Koshel *et al.,* 2014). All the clade support valuesare<50 showing the high quality of the tree. Individual gene trees drawn by using the nucleotide data of 18S rRNA, 28S rRNA, D3-expansion domain, ITS1-5.8S-ITS2 and protein data of COX1 and ND1 also resulted in similar tree only with slight variations (Figure 6). Highest variation in individual tree from MLST tree was shown by ITS1-5.8S-ITS2 and ND1 further supporting the polymorphism results of highest variability.

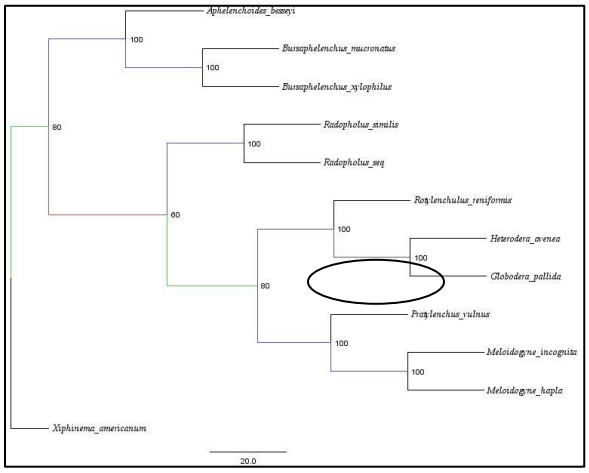


Figure 5: Unrooted phylogenetic tree obtained from multi-locus sequence data of our sequence data with 12 PPNs sequence data

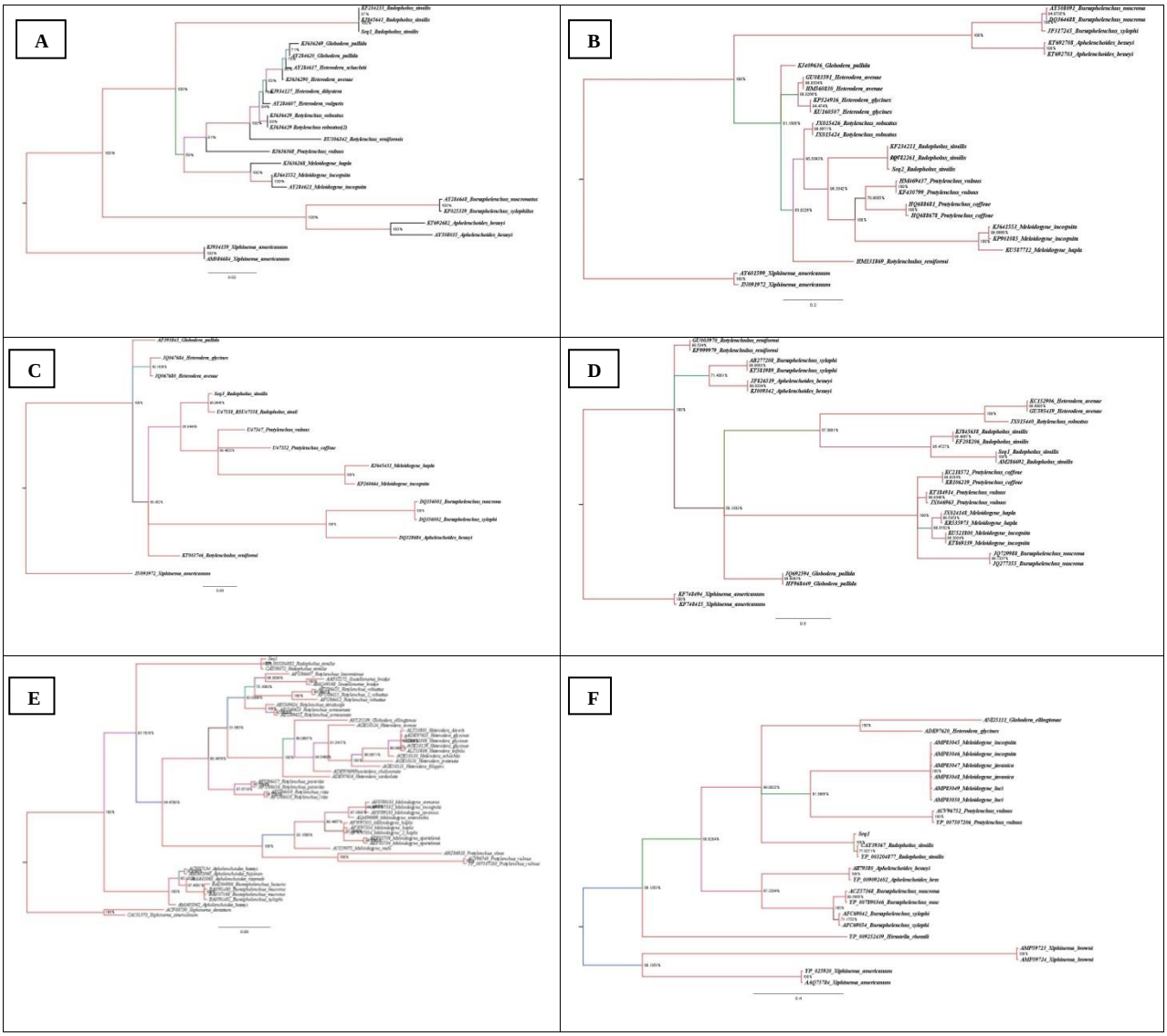


Figure 6: Unrooted individual trees of A) 18S rRNA, B) 28S rRNA, C) D3-expansion domain, D) ITS1-5.8S-ITS2, E) COX1 and F) ND1obtained using MrBayes-3.2.6.

1.5. Intra-species variation analysis

From the 93 taxa used for the study a total number of 35 haplotypes were obtained. The ITS region from India present in NCBI (Accession No.- AM286692) only contained ITS1 and 5.8S region and therefore was not used in this analysis as missing data can have a large effect on the outcome of a network analysis (Tan *et al.,* 2010). When this sequence was compared with the ITS regions of *R. similis* sequenced during this study, it showed 100% identity. Other ITS regions from India (Accession No.- MF197892-MF197883) were also not used in this study as all the sequences where from almost same regions (Kerala and Karnataka, India) as the sequence in this study. The sequence analysis done using BLAST and MSA among the population from India shows more than 99% identity (Supplementary Table S4 and S5).

Haplotype distribution study, divided the 93 sequences into 35 different haplotypes with a Haplotype diversity (Hd) of 0.8534. In nematodes, generally the ITS region had an overall higher level of haplotype diversity than COI, but lower nucleotide diversity as recently studied in the case of *Aphelenchoides besseyi* population (Xu *et al.,* 2020). The haplotype diversity and nucleotide diversity of ITS regions are given in Table 11. Further phylogenetic network analysis suggested that the 93 ITS genotypes can be split into two groups with frequent gene flow interactions (Figure 7). The majority of haplotypes were closely related by 1–3 steps in the TCS network analysis, while Hap17 from India was more divergent from other haplotypes.

Table 11: Sequence statistics and diversity indices computed for ITS genes of the Indian isolate along with *R. similis* from various geographical regions

Alignment length (bp)	**613**
Number of haplotypes	35
Variable sites	333
Singletons Variable sites	266
Parsimony informative sites	67
Nucleotide diversity	0.031
Haplotype diversity	0.853

Among the 35 haplotypes 27 haplotypes consisted of just one sequence. The maximum number (34) of sequences where present in hap 4, which interestingly contained sequences from all 5 continents. The next largest haplotype contained 7 sequences, among which one was from Belgium and six sequences from Sudan. Fascinatingly, sequences from same geographical regions were spread to different haplotypes, except for Australian population. All the 10 Australian population were present in the same haplotype, Hap4.

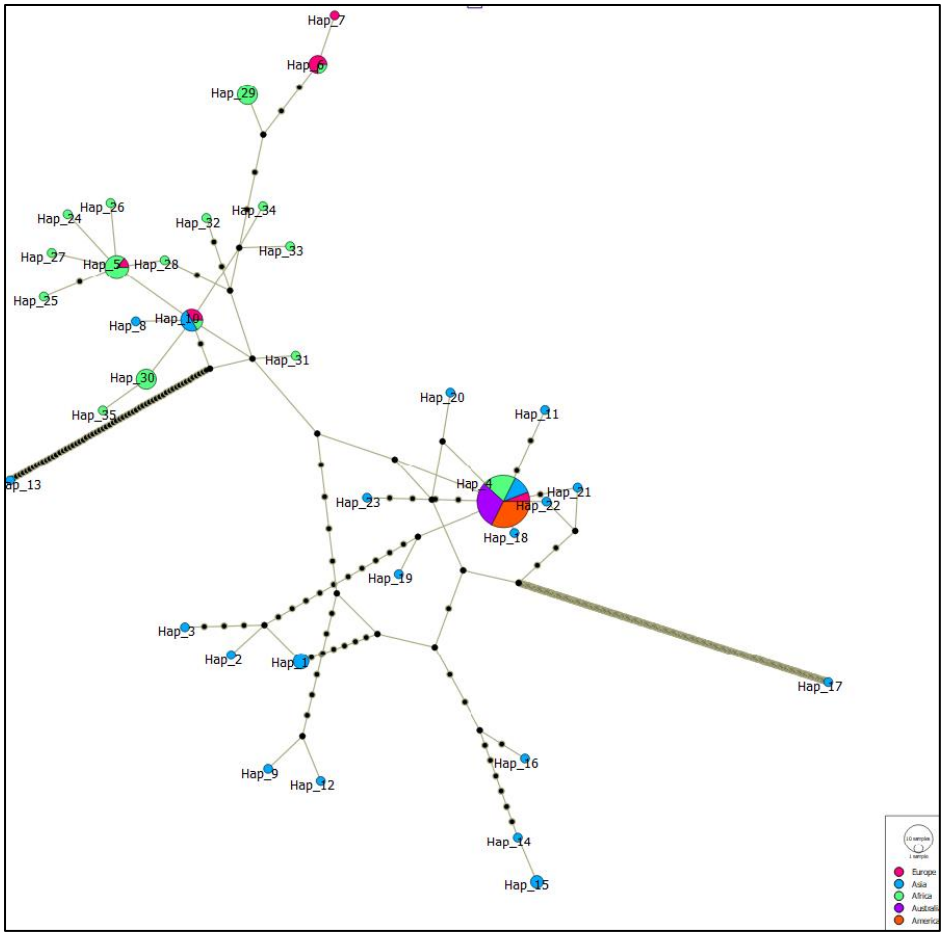


Figure 7: Haplotype network analysis estimated using TCS networks

Phylogenetic relationships of *R. similis* R01 with other *R. similis* as inferred from the partial sequences of ITS1-5.8S-ITS2 regions is given in Figure 8. Bayesian inference and maximum likelihood analyses gave rise to similar trees with high support values for most branches. The trees generated by both methods agreed with the result obtained by haplotype network analysis. The clade with the largest number of taxa contained all the Australian species and was made up of mostly the members present in haplotype group AUS. The Indian population of *R. similis* R01 formed a clade with one of the Chinese species. As indicated by haplotype network, phylogenetic analysis also found the Indian population to be highly variable as compared to species from other populations. Other than Australian population several other species from same geographical regions like Malaysia, Colombia and Costa Rica were part of same clade. Whereas species from geographical areas like China, Belgium and Sudan were part of several different clade.

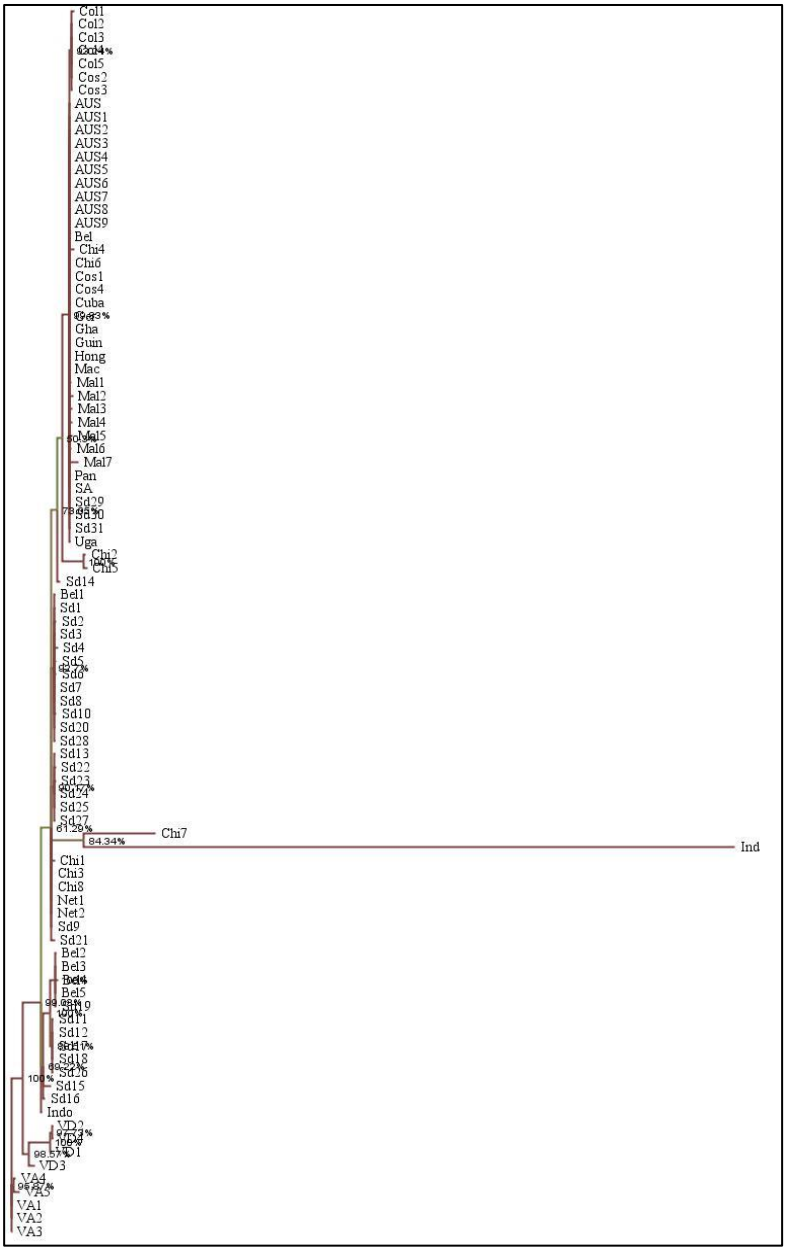


Figure 8: Unrooted tree of ITS1-5.8S-ITS2 regions of showing phylogenetic relation within *R. similis* from various geographical regions

2. TRANSCRPITOME SEQUENCING, ASSEMBLY AND ANNOTATION OF *R. SIMILIS*

2.1. Illumina Hi-seq transcriptome sequencing, assembly and annotation

The Illumina Hi-seq sequencing run resulted in a total 10,796,325 number of reads for both forward and reverse sequences with an average sequence length of 101 nucleotides and an average GC content of 52%. After trimming to remove adaptors, low quality reads with ambiguous bases and polyN tail sequences, 10,761,831 clean reads remained with a mean length below 101 bp. Picard was used to find and remove the PCR duplicates present in the raw data. A total of 10,032,884 reads were obtained after

duplicates removal. Using Trinity, raw data, trimmed data and PCR duplicate removed data were assembled with a minimum Kmer coverage of 25 to produce a collection of contigs representing the pool of cDNA fragments generated in this study (Table 12). The quality of the assembly generated with each data was assessed using BUSCO before subsequent analysis. Since the assembly quality was almost similar in all three methods (Figure 9), therefore raw data assembly was used for further analysis since it guarantees minimum information loss.

Table 12: Statistics of assembly done on raw data, trimmed data and PCR duplicate removed data

	Number of contigs	Contig N50	Shortest contig	Longest contig	Mean contig size	Percent GC
Raw reads	85253	2420	201	10747	1311	51.09
Filtered	74854	2300	201	10747	1274.10	50.98
Duplicates removed	82893	2443	201	12195	1329.78	51.04

After assembly, the overall redundancy was reduced by removal of transcripts wholly represented in other transcripts at 98% similarity using CD HIT-EST. A total of 62312 unigenes (73.09%) were obtained ranging in size form 201– 10747bp with a mean contig size of 1046 bp. Coding regions present within the unigenes were predicted using the tool Transdecoder generating a total of 50685 cds which are further used for annotation.

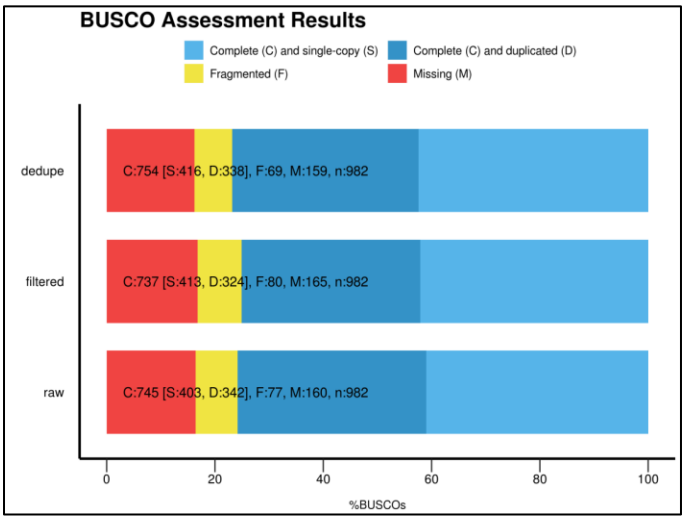


Figure 9: BUSCO assembly assessment results showing the comparison of assembly using raw data, trimmed data and PCR duplicate removed data

The annotation of the predicted cds resulted in 37,938 annotated sequences (74.85%). Gene Ontology identifiers were also searched for these cds using Blast2Go and a total of 34,105 cds were assigned at least one GO term (Figure 13). The predominant terms for 'biological process' were 'cellular process' and 'metabolic process' (30% and 29%, respectively), for 'cellular component' were 'cell part' and 'cell' (20% and 20%, respectively), and for 'molecular function' were 'binding' and 'catalytic activity' (51% and 40%, respectively) (Figure 10, 11, 12).

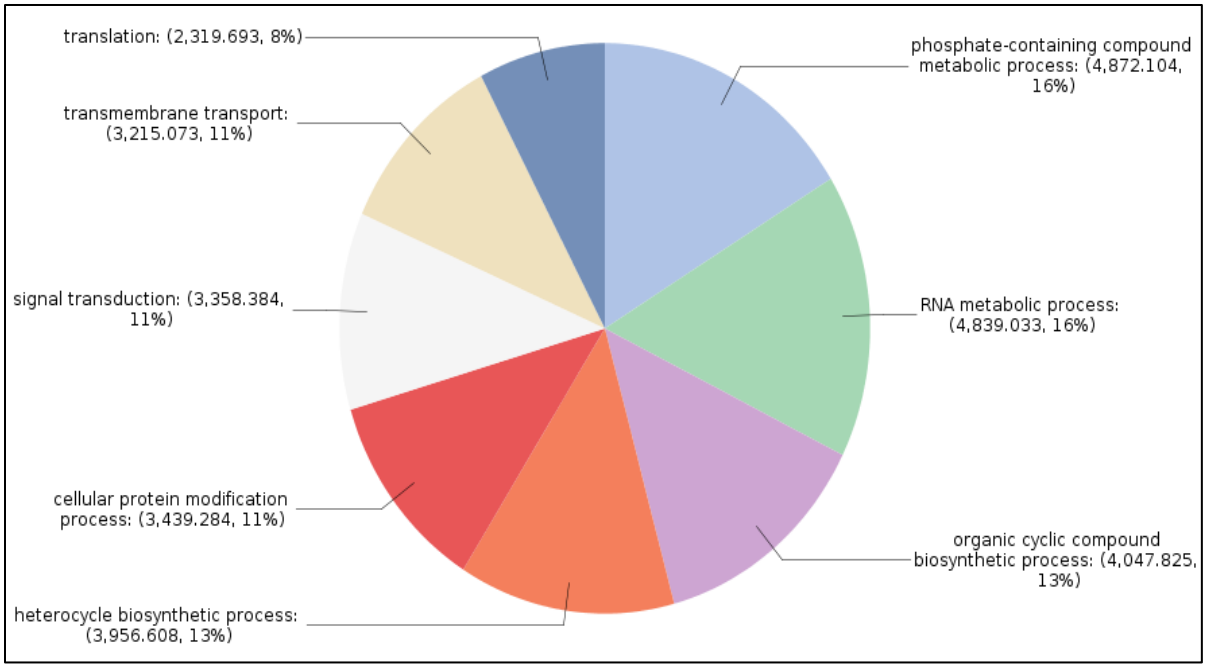

translation: (2,319.693, 8%)
phosphate-containing compound metabolic process: (4,872.104, 16%)
transmembrane transport: (3,215.073, 11%)
signal transduction: (3,358.384, 11%)
RNA metabolic process: (4,839.033, 16%)
cellular protein modification process: (3,439.284, 11%)
organic cyclic compound biosynthetic process: (4,047.825, 13%)
heterocycle biosynthetic process: (3,956.608, 13%)

Figure 10:Summary of biological processes associated with the transcriptome of *R. similis*

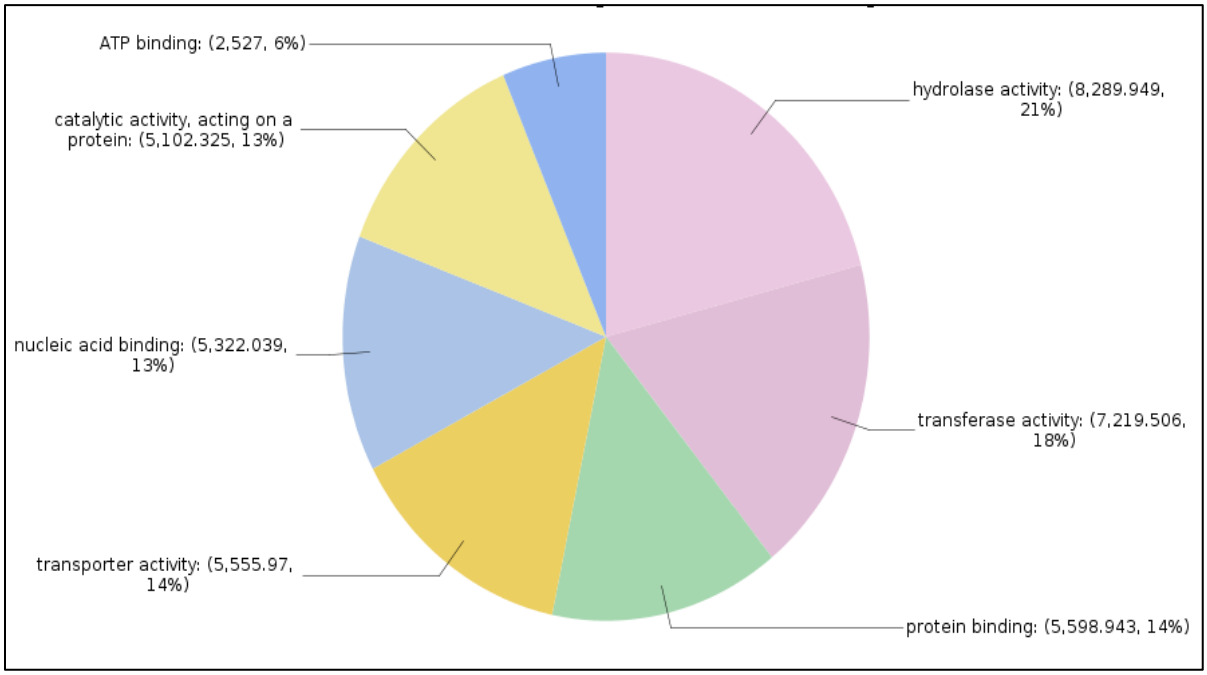

ATP binding: (2,527, 6%)
hydrolase activity: (8,289.949, 21%)
catalytic activity, acting on a protein: (5,102.325, 13%)
nucleic acid binding: (5,322.039, 13%)
transferase activity: (7,219.506, 18%)
transporter activity: (5,555.97, 14%)
protein binding: (5,598.943, 14%)

Figure 11: Summary of molecular function associated with the transcriptome of *R. similis*

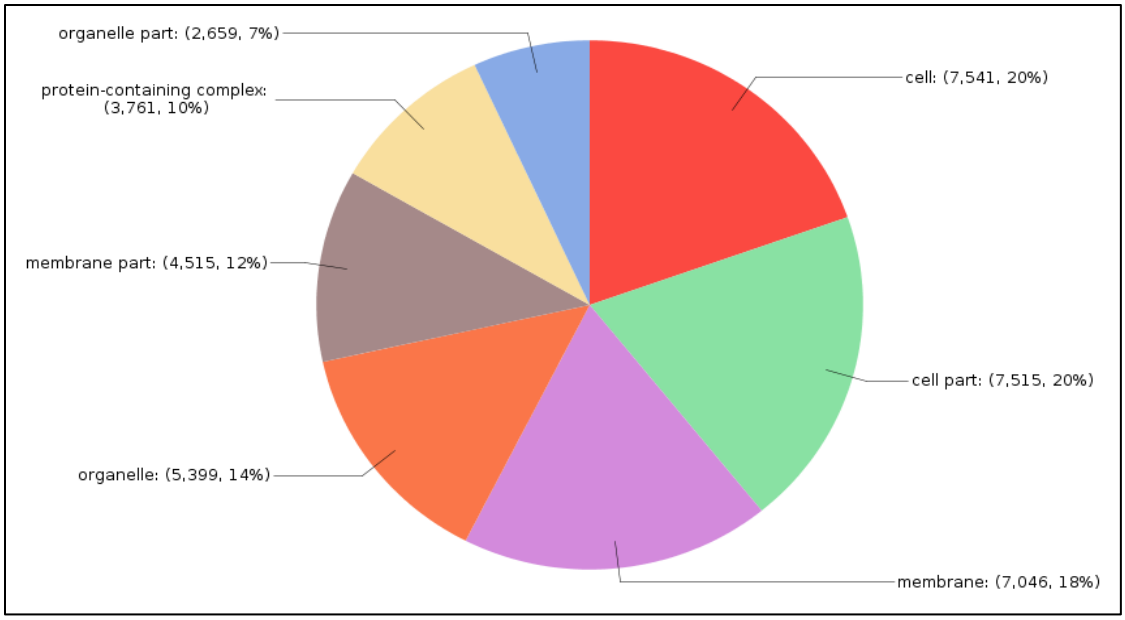


Figure 12: Summary of cellular components associated with the transcriptome of *R. similis*

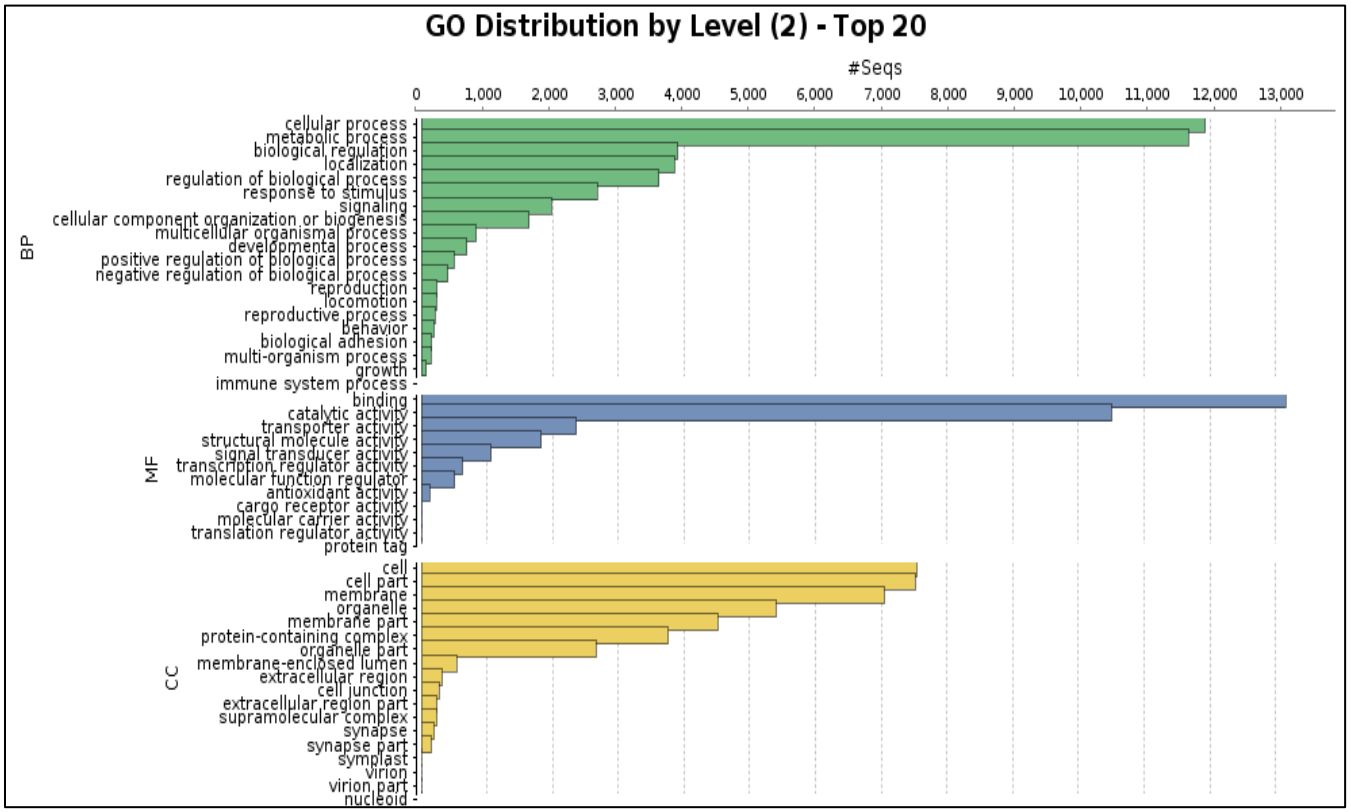


Figure 13: Gene ontology annotation results of *R. similis* transcriptome data where BP represents biological processes, MF represents molecular function and CC represents cellular components

Proteins are composed of domains, recurrent protein fragments with distinct structure, function, and evolutionary history.Domain repeats are often involved in interactions with proteins or other ligands such as DNA or RNA. The transcriptome of *R. similis* contained several important domains and repeats

aspredicted by interproscan (Figure14 and 15). WD40 was the most abundant repeat region present within the transcriptome whereas protein kinase domain was the most common domain present.

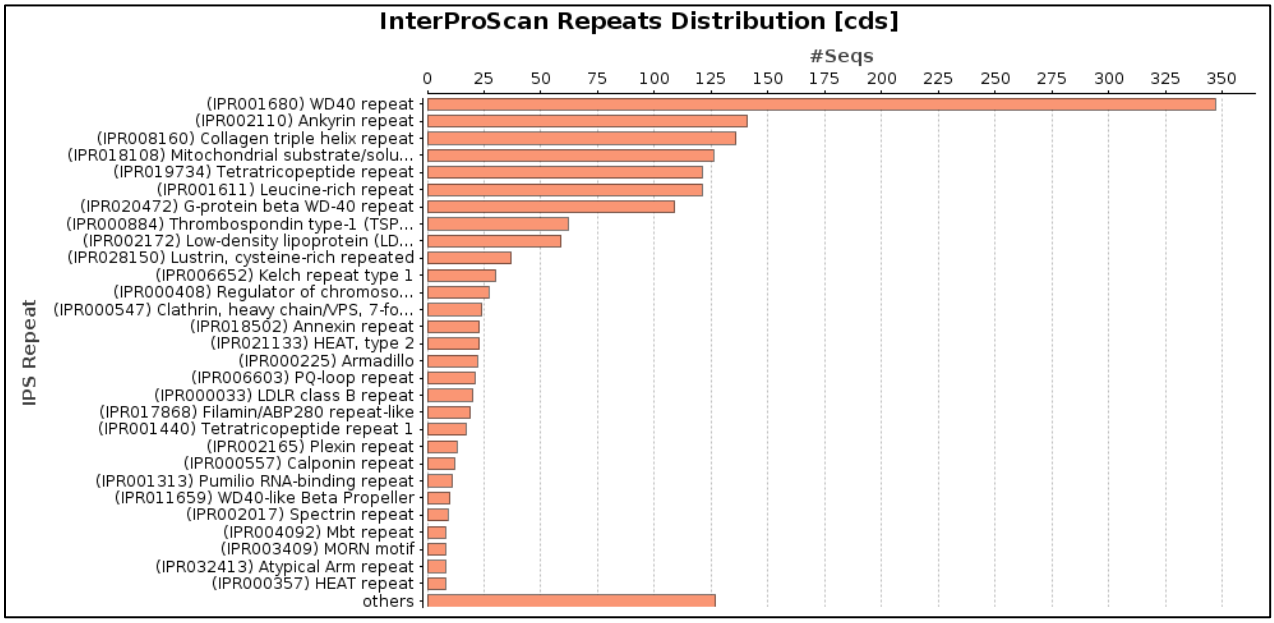


Figure 14: Repeats predicted in the *R. similis* transcriptome data using InterproScan

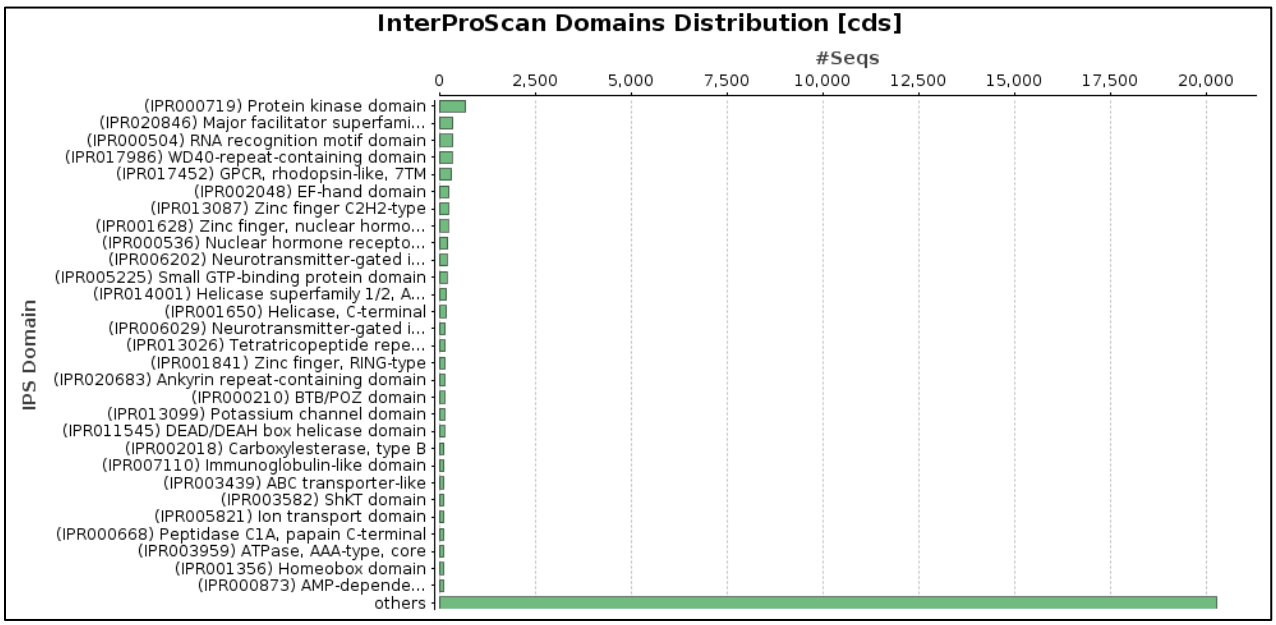


Figure 15:Domains predicted in the *R. similis* transcriptome data using InterproScan

Perl script MISA was used to identify SSRs and generated 2,310 SSRs found in 2,074 transcripts, with a frequency of one SSR per 1.94 kb of the sequence (Table 13). Tri-nucleotide SSRs represented the largest fraction (80.73%) followed by mononucleotide (10.69%) and di-nucleotide (8.05%) repeats. Only a few tetra- (9) and hexa-nucleotide repeat (3) SSRs were identified in the *R. similis* transcripts.

Table 13. Details of the expressed Simple Sequence Repeats identified in *R. similis* transcripts

SSR mining	
Total number of sequences examined	50685
Total size of examined sequences (bp)	44964393
Total number of identified SSRs	2310
Number of SSR containing sequences	2074
Number of sequences containing more than one SSR	198
Number of SSRs present in compound formation	99
Frequency of SSRs	One per 1.94 kb

2.2. ES protein prediction from the transcriptome data set of *R. similis*

Of the 50685 proteins encoded in the *R. similis* transcriptome, 3,584 proteins from SignalP and 5,156 proteins from Phobius were predicted as secretory proteins. The results from these two programs were combined and yielded a total of 6,147 proteins as classical secretory proteins. The classical secretory proteins obtained were then analyzed by TMHMM and 2,307 proteins were predicted as transmembrane proteins. The remaining 3840 proteins when checked by TargetP, predicted 230 proteins as mitochondrial proteins at 95% specificity, which were then removed from the secretome dataset. The remaining 3,610 proteins were queried with PredGPI Predictor yielding 133 potential GPI anchored proteins. The application of ProtComp v9.0 to the remaining 3,477 ORFs yielded a total of 1116 ORFs (40%) as ES proteins.

To examine the length distribution of predicted ES proteins of *R. similis*, all the 1116 ES proteins were analyzed using bioinformatics tools. The longest protein was 988 amino acid (aa) and the shortest one was 100 aa. The length of most secretory proteins (76.67% of the total identified proteins with a signal peptide) was between 100 and 300 aa (Figure 16). Within this group, about 56.36% were 100-200 aa long, suggesting that most secretory proteins probably fall in the shorter length range (Jing *et al.,* 2017).

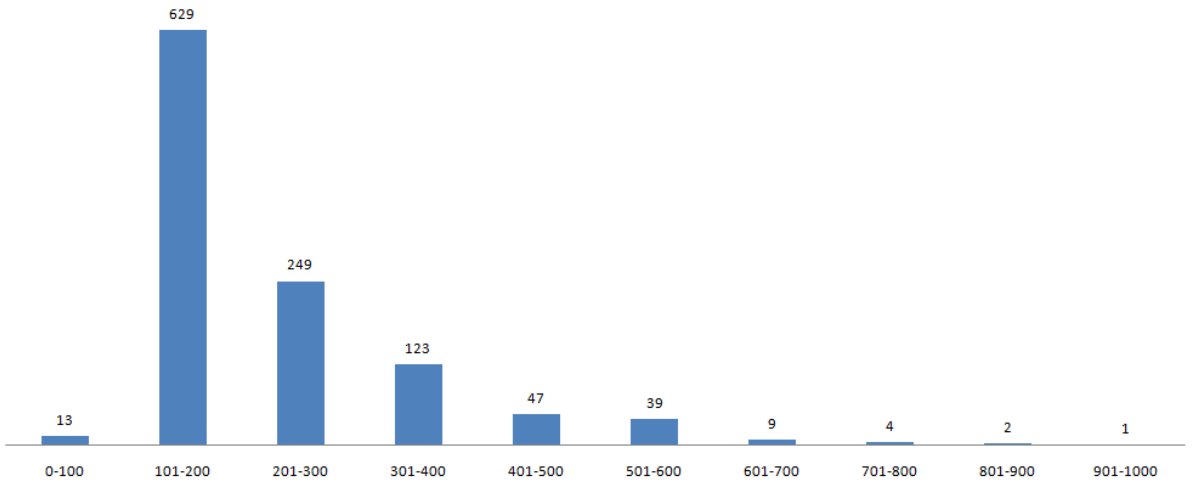


Figure 16: Length distribution of secretory proteins of *R. similis*

2.3. Functional and comparative annotation of ES proteins of *R. similis*

ES proteins were functionally categorized into Biological Process, Molecular Function and Cellular Component categories with Gene Ontology terms. Out of 1116 ES proteins, 404 (36.2%) proteins were assigned to at least a GO terms using InterproScan. The querying of 1116 secretory proteins against the COG database was performed for functional classification. A total of 125 proteins could be assigned to the COG classification, of which 30 (24%) potentially participated in post-translational modifications, protein turnover, and chaperones followed by 22% involved in cell wall/membrane/envelope biogenesis(Figure 17). Proteins participating in inorganic ion transport and metabolism; carbohydrate transport and metabolism; transcription; lipid transport and metabolism accounted for about 9%, 8%, 6%, 5%, respectively. Blasting PHI yielded a total of 1098 secretory proteins that could be connected to pathogenicity (Figure 18). Of these, 565 ES proteins had a blast hit and thus assigned a function. ES proteins that were present in the PHI category of lethal and loss of function are listed in Table 14.

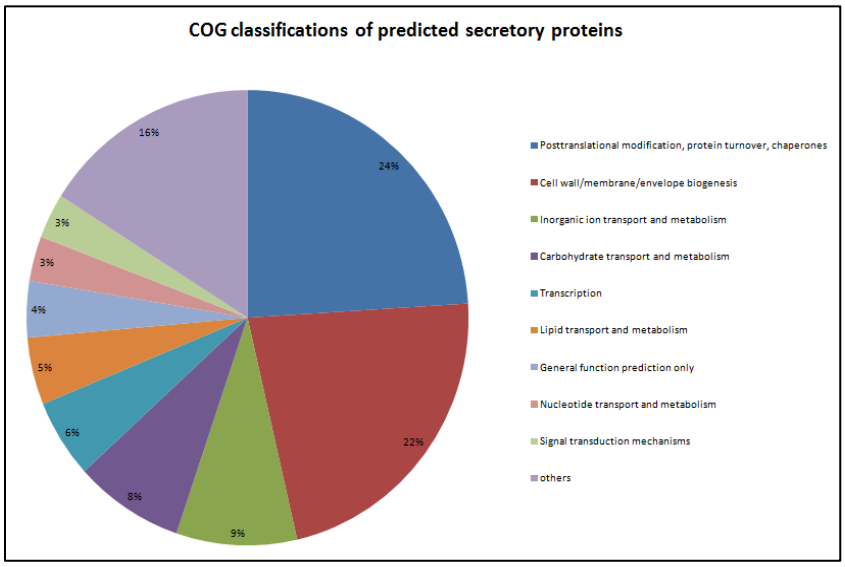


Figure 17: Summary of the secretome of *R. similis* having a hit against COG of proteins database

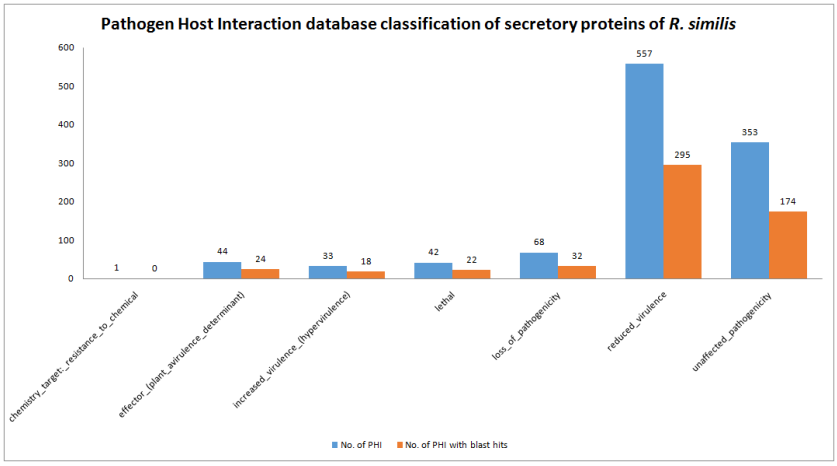


Fig. 18: Summary of the secretome of *R. similis* having a hit against Pathogen Host Interaction database

Table 14: Functional classes of the secretory proteins of *R. similis* present in PHI category; lethal and loss of pathogenicity

Classification	Contig ID	Function
lethal	DN15507_0_1_1_52316	brix domain-containing protein
	DN22281_2_4_3_96594	CRE-MLT-8 protein
	DN20581_0_1_4_80588	disintegrin and metalloproteinase domain-containing protein 10-like isoform X3
	DN20581_0_1_7_80592	disintegrin and metalloproteinase domain-containing protein 10-like isoform X3
	DN13649_0_1_2_52687	dual specificity protein kinase TTK
	DN17591_0_1_3_33660	FMRF amide-like protein 7
	DN21270_0_1_1_106986	Hedgehog protein, Hint domain and Hint domain C-terminal and Hint domain N-terminal-containing protein
	DN17223_0_1_2_32341	Homolog of Odr-2 (Two)
	DN17223_0_1_8_32357	Homolog of Odr-2 (Two)
	DN799_0_1_1_51840	hypothetical protein ANCCAN_04955
	DN32762_0_1_1_35459	photosystem I reaction center subunit N, chloroplastic
	DN20905_0_1_21_60920	putative amphid protein
	DN20905_0_1_23_60922	putative amphid protein
	DN20905_0_1_5_60918	putative amphid protein
	DN16832_0_1_1_21379	putative effector protein
	DN3434_0_1_1_109742	Spermatogenesis-associated protein 5
	DN2554_0_1_1_81914	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
	DN22193_0_1_5_89490	TKL/MLK/MLK protein kinase
	DN18989_0_1_1_76347	Protein CBG04378
	DN18989_0_1_4_76353	Protein CBG04378
	DN18989_0_1_8_76361	Protein CBG04378
	DN18989_0_1_9_76364	Protein CBG04378
Loss of pathogenicity	DN21531_0_1_10_15795	Bm4501, isoform a
	DN21531_0_1_29_15863	Bm4501, isoform a
	DN16152_0_1_1_105702	common central domain of tyrosinase
	DN20509_0_1_5_81582	DB module
	DN20509_0_1_7_81589	DB module
	DN20509_0_1_8_81591	DB module
	DN18038_0_1_1_10167	DEgenerin Like
	DN19785_0_2_5_71430	EGF-like domain protein
	DN19785_0_2_8_71451	EGF-like domain protein
	DN20822_0_2_3_101157	extracellular solute-binding protein
	DN14824_0_1_1_37987	glycoside hydrolase
	DN16655_0_2_1_66522	hypothetical protein BE17_10025
	DN16655_0_2_2_66526	hypothetical protein BE17_10025
	DN17781_0_1_2_32629	Hypothetical protein SRAE_2000371800
	DN800_0_1_1_62565	hypothetical protein Tcan_12799, partial
	DN21180_0_2_1_13339	hypothetical protein Tcan_16445
	DN22252_1_1_6_97964	Kunitz/Bovine pancreatic trypsin inhibitor domain protein
	DN18181_0_1_1_100088	PAN domain protein
	DN18181_0_1_2_100093	PAN domain protein

	DN38095_0_1_1_65254	PAN-1 domain and Apple-like domain-containing protein
	DN18499_0_2_1_72506	Poly(U)-specific endoribonuclease
	DN15097_0_1_2_75338	Protein C39E9.8, isoform d
	DN20566_0_1_12_80798	Protein CBR-COL-127
	DN20566_0_1_20_80823	Protein CBR-COL-127
	DN18946_0_1_6_76378	Protein kinase domain and Serine/threonine-/dual specificity protein kinase, catalytic domain and Protein kinase-like domain-containing protein
	DN18478_0_1_2_72234	saposin precursor
	DN42523_0_1_1_65668	suppressor of tumorigenicity 14 protein-like isoform X6
	DN885_0_1_1_62563	surfactant protein B
	DN28218_0_1_1_70361	hypothetical protein Tcan_07957
	DN22638_0_1_1_94807	Peptidase M12A domain containing protein
	DN15574_0_1_3_52264	Protein CBG13414
	DN22668_0_1_1_94818	mammalian ependymin-related protein 1

Of the 1116 ES proteins, we could annotate a total of 477 (42.74%) proteins in 113 different protein domains and families. The 30 most represented Interpro domains are shown in Figure 19. which included ShKT domain, Peptidase C1A, papain C-terminal, C-type lectin-like, Pancreatic trypsin inhibitor Kunitz domain, Immunoglobin-like domain, EB domain, EF-hand domain and Glycosyl hydrolase family 30, TIM-barrel domain. The ShKT domain, also called NC6 (nematode six-cysteine) domain,is a large superfamily of proteins that contains domains resembling the two toxins BgK and ShK. Interestingly, C-type lectin are involved in a variety of functions including cell-cell adhesion, immune response to pathogens and apoptosis.

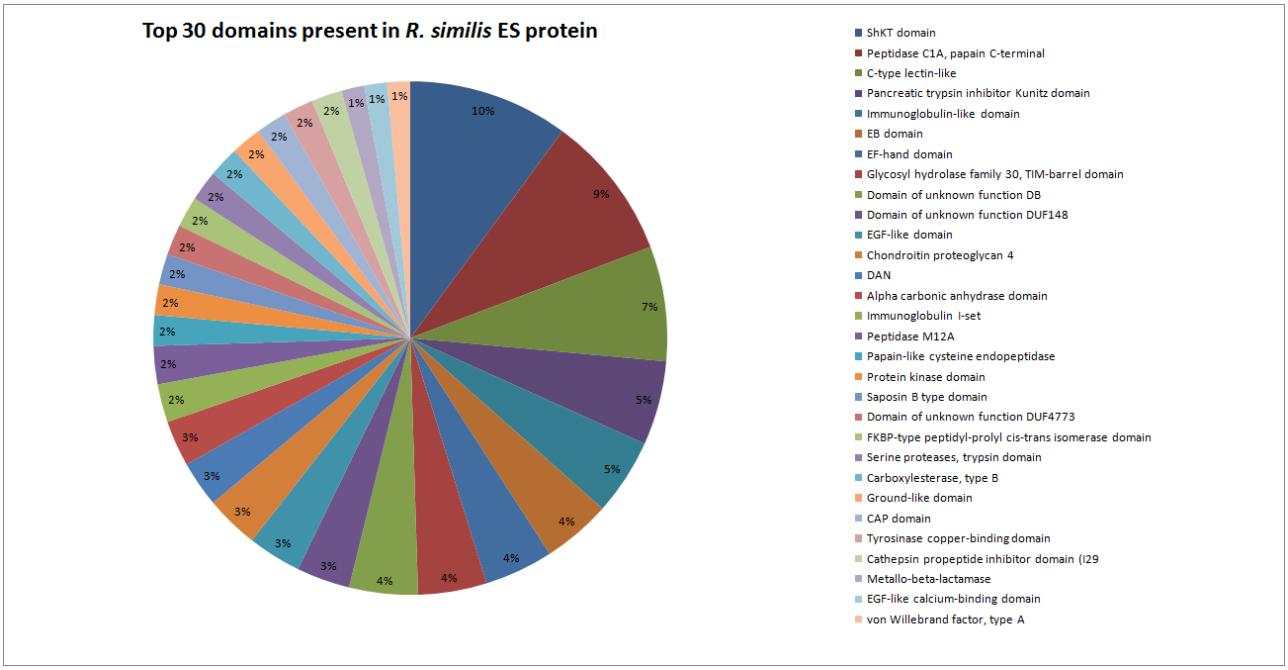


Figure 19: Summary of the domains predicted in the secretome of *R. similis*

A total of 119 (11.8%) proteins were mapped to 57 KEGG pathways. The 20 most represented KEGG pathways included protein processing in lysosome, metabolic pathways, glycan degradation, nitrogen metabolism, cAMP signaling pathway, MAPK signaling pathway, PI3K-Akt signaling pathway and ubiquitin mediated proteolysis (Table 15). The most represented KEGG BRITE objects mapped to ES

proteins included enzymes, exosomes, chaperons, transcription factors, spliceosomes, protein kinases, ribosomes and peptidases.

Table 15: Top biological pathways identified in *R. similis* ES proteins

Contigs IDs	Pathways involved
DN21695_0_1_10_22092 ,DN22160_1_1_3_90414 ,DN21695_0_1_5_22072, DN22160_1_1_2_90349	**AGE-RAGE signaling pathway in diabetic complications**
DN40645_0_1_1_75551 ,DN22012_1_3_11_54904 ,DN32558_0_1_1_72070	**Arachidonic acid metabolism**
DN13448_0_1_1_75661 ,DN22160_1_1_3_90414 ,DN22160_1_1_2_90349	**EGF receptor signaling pathway**
DN22160_1_1_3_90414 ,DN22308_1_5_1_92788 ,DN22160_1_1_2_90349	**Endothelin signaling pathway**
DN13448_0_1_1_75661 ,DN22160_1_1_3_90414 ,DN22160_1_1_2_90349	**FGF signaling pathway**
DN22160_1_1_3_90414 ,DN22160_1_1_2_90349	**Hypoxia response via HIF activation**
DN21695_0_1_10_22092 ,DN22160_1_1_3_90414 ,DN21695_0_1_5_22072, DN22160_1_1_2_90349	**Inositol phosphate metabolism**
DN33250_0_1_1_37733 ,DN15388_0_1_1_78432 ,DN20756_0_1_24_20727, DN11184_0_1_1_8602 ,DN20960_0_1_1_60504 ,DN4684_0_1_1_109952, DN20756_0_1_6_20648 ,DN22025_1_3_2_55094 ,DN6564_0_1_2_92477, DN18723_0_1_2_94145 ,DN22156_0_2_5_91281 ,DN22025_1_3_8_55116, DN20756_0_1_5_20638 ,DN20756_0_1_7_20658 ,DN18042_0_1_3_10114, DN20756_0_1_2_20621 ,DN12734_0_1_2_86054 ,DN20885_0_1_16_102032 , DN7155_0_1_2_33256 ,DN20056_1_2_1_61361 ,DN20056_1_2_7_61372, DN18478_0_1_2_72234 ,DN20756_0_1_19_20691 ,DN20056_1_2_2_61367, DN20756_0_1_22_20713 ,DN20756_0_1_17_20682 ,DN20885_0_1_21_102057, DN7155_0_1_1_33253 ,DN33153_0_1_1_79585 ,DN22156_0_5_2_91249, DN22025_1_3_9_55120 ,DN22156_0_2_3_91247 ,DN22025_1_3_4_55105, DN12734_0_1_1_86052	**Lysosome**
DN6262_0_1_1_83427 ,DN20756_0_1_24_20727 ,DN20960_0_1_1_60504, DN20885_0_1_21_102057 ,DN20756_0_1_6_20648 ,DN21358_2_1_11_67898, DN18432_0_2_1_72647 ,DN32558_0_1_1_72070 ,DN21824_2_2_13_31601, DN20756_0_1_5_20638 ,DN20756_0_1_7_20658 ,DN18629_0_1_5_57947, N20756_0_1_2_20621 ,DN20885_0_1_16_102032 ,DN15114_0_1_2_11073, DN18042_0_1_3_10114 ,DN20756_0_1_19_20691 ,DN20756_0_1_22_20713, DN18629_0_1_2_57940 ,DN20756_0_1_17_20682 ,DN21695_0_1_5_22072, DN18432_0_2_2_72655 ,DN21824_2_2_22_31621 ,DN21695_0_1_10_22092	**Metabolic pathways**
DN20960_0_1_1_60504 ,DN22160_1_1_3_90414 ,DN22160_1_1_2_90349	**mTOR signaling pathway**
DN20759_0_1_7_19843 ,DN20759_0_1_27_19904 ,DN20759_0_1_24_19899, DN20759_0_1_35_19945 ,DN20759_0_1_5_19835 ,DN20759_0_1_34_19941	**Nitrogen metabolism**
DN20756_0_1_6_20648 ,DN20756_0_1_19_20691 ,DN20756_0_1_22_20713, DN20756_0_1_2_20621 ,DN20756_0_1_17_20682 ,DN20885_0_1_16_102032, DN20885_0_1_21_102057 ,DN20756_0_1_24_20727 ,DN20756_0_1_5_20638, DN20756_0_1_7_20658 ,DN18042_0_1_3_10114	**Other glycan degradation**
DN18723_0_1_2_94145 ,DN11184_0_1_1_8602 ,DN33153_0_1_1_79585, DN20960_0_1_1_60504	**Phagosome**
DN21695_0_1_10_22092 ,DN22160_1_1_3_90414 ,DN21695_0_1_5_22072, DN22160_1_1_2_90349	**Phosphatidylinositol signaling system**
DN23035_0_1_1_14680 ,DN3537_0_1_2_93520 ,DN24053_0_1_1_104743	**PI3 kinase pathway**
DN17735_0_1_1_32707 ,DN18432_0_2_1_72647 ,DN18629_0_1_2_57940, DN18432_0_2_2_72655 ,DN22308_1_5_1_92788 ,DN18629_0_1_5_57947	**Purine metabolism**
DN17735_0_1_1_32707 ,DN18629_0_1_5_57947 ,DN18629_0_1_2_57940	**Pyrimidine metabolism**
DN20756_0_1_6_20648 ,DN20756_0_1_19_20691 ,DN20756_0_1_22_20713, DN20756_0_1_2_20621 ,DN20756_0_1_17_20682 ,DN20885_0_1_16_102032, DN20885_0_1_21_102057 ,DN20756_0_1_24_20727 ,DN20756_0_1_5_20638, DN20756_0_1_7_20658	**Sphingolipid metabolism**
DN13448_0_1_1_75661 ,DN17057_0_1_3_44294 ,DN17057_0_1_2_44290	**Wnt signaling pathway**

Plant parasitic nematodes produce a variety of carbohydrate activity enzymes (CAZymes) for the degradation of plant cell wall to facilitate infection and gain nutrition. In total, 162 (8.5%) ES proteins were mapped to 38 CAZyme families. The 10 most represented CAZyme families which included GT1 (alpha-sialyltransferase), GH25 (lysozyme), CBM50 (Chitinase), CBM13 (xylanase) and GH30_1 (endo-b-1,4- xylanase) are given in Table 16.

Table 16: Top 10 most represented CAZyme families in *R. similis* ES proteins

CAZyme familly	Functional activity	Number of secreted proteins
GT1	alpha-sialyltransferase (EC 2.4.99.-)	16
GH30_1	endo-beta-1,4-xylanase (EC 3.2.1.8)	10
CBM13	xylanase	9
GH25	lysozyme (EC 3.2.1.17)	8
CBM2	cellulose-binding function.	5
CBM50	Chitinase	4
GT8	beta-galactoside alpha-2,6-sialyltransferase (EC 2.4.99.1)	3
GH43_35	beta-xylosidase (EC 3.2.1.37)	2
GH5	beta-agarase (EC 3.2.1.81).	2
GT22	Dol-P-Man: Man6GlcNAc2-PP-Dol alpha-1,2-mannosyltransferase (EC 2.4.1.259)	2

The ES protein sets of *R. similis* were compared against those of 32 other species.Nine plant parasitic nematodes other than *R. similis* were used for the study. Among the selected species, several parasitic species other than nematodes were also included - two plant pathogenic fungi (*Magnaporthe grisea* and *Fusarium graminearum*), one nematode parasite of animals (*Brugia malayi*) and two insects that feed on living plant tissue (*Acyrthosiphon pisum* and *Bombyx mori*). Eighteen other species were listed as blacklisted/non-target species:

- Insects: *Apis mellifera, Drosophila melanogaster*
- Plants: *Arabidopsis thaliana, Citrus sinensis, Musa acuminata*
- Free living nematodes: *Caenorhabditis briggsae, Caenorhabditis elegans, Pristionchus pacificus*
- Fungi: *Saccharomyces cerevisiae, Schizosaccharomyces pombe*
- Vertebrates: *Bos taurus, Danio rerio, Gallus gallus, Homo sapiens, Mus musculus, Sus scrofa, Tetraodon nigroviridis, Xenopus tropicalis*

Around 109 sequences shared homology with one or more blacklisted/non-target species and thus were removed. The ES proteins non-homologous to the blacklisted/non-target proteome were further screened for RNAi phenotypes in *C. elegans.* Around 170 ES proteins had one or more hits with RNAi phenotypes in *C. elegans* (Table 17).

Table 17:Top 5 most represented RNAi phenotypes in *R. similis* ES proteins

ES protein ID	C.elegans gene ID	RNAi phenotype
DN20465_0_1_11_8905, DN20465_0_1_12_8913, DN20465_0_1_21_8928, DN20465_0_1_28_8963, DN19676_0_2_1_77954, DN17486_0_1_4_15081, DN4914_0_1_2_83857, DN22202_1_2_4_95315, DN4914_0_1_1_83856, DN7506_0_1_2_79730, DN14013_0_1_2_53059, DN13380_0_1_4_100547, DN22156_0_2_3_91247, DN22156_0_2_5_91281, DN18424_0_1_2_72199, DN18424_0_1_5_72204	R186.1, C13C12.1b, Y55F3AM.10, T26C5.5, Y37E11B.4, Y39B6A.23, C35D10.5, T07C4.11a, C52E4.1, Y73B6BL.21a	Embryonic lethal
DN20161_1_2_2_25687, DN908_0_1_2_47690, DN908_0_1_3_47696, DN17617_0_1_2_66844, DN17617_0_1_4_66850, DN17617_0_1_3_66847,	ZC84.3, Y48A6A.1, F56B6.4, C13G3.3, ZK973.6, F57F5.1	Larval arrest

DN13448_0_1_1_75661, DN22961_0_1_1_75170, DN22156_0_5_2_91249		
DN12655_0_1_1_106130, DN22193_0_1_38_89621, DN18504_0_1_6_102409, DN6262_0_1_1_83427, DN16599_0_1_2_93931	F08C6.1c, CC4.2, R12A1.4, Y47D3B.10, Y59A8B.20	Locomotion variant
DN21824_2_2_22_31621, DN21824_2_2_13_31601, DN16152_0_1_1_105702, DN12016_0_1_1_50251, DN12016_0_1_2_50254, DN19853_0_1_2_58348	K08E3.1, F21C3.2, Y54G2A.23	Germ cell ionizing radiation response variant
DN6564_0_1_2_92477, DN22047_1_3_3_54435, DN4684_0_1_1_109952, DN22025_1_3_2_55094, DN22025_1_3_4_55105, DN22025_1_3_9_55120, DN20056_1_2_7_61372, DN33250_0_1_1_37733	T11A5.6c, C25B8.3c, C25B8.3b, H22K11.1	Hypoosmotic shock hypersensitive, lethal

3. IDENTIFICATION AND SEQUENCING OF NOVEL TARGET GENES

3.1. Sequence data analysis and NCBI submission

The sequenced target genes were filtered, trimmed and assembled to produce high quality sequences. NCBI Blast similarity search was done for all the sequenced genes to validate the PCR product and the sequencing result. The percentage identity and similarity of the amplified target gene obtained from BLAST search is given in Table 18. The assembled sequences confirmed using BLAST similarity search were then submitted to NCBI-GenBank. The NCBI accession number, nucleotide length and NCBI description of each target gene is provided in Table 19.

Table 18: Results of NCBI Blast similarity search for nine sequenced target genes

Target id	Length	BLAST hit	Identity (%)	Description [Hit organism]
Rs-AB	1104	EJW78952	78.69	abhydrolase domain-containing protein FAM108C1 [*Wuchereria bancrofti*]
Rs-Ac16	249	ABD98404	68.57	immunodominant hypodermal antigen Ac16 [*Ancylostoma caninum*]
Rs-AsLM	996	XP_024506364	45.49	Astacin-like metalloendopeptidase [*Strongyloides ratti*]
Rs-CDh	354	PIO71599	56.07	carbonate dehydratase, eukaryotic-type [*Teladorsagia circumcincta*]
Rs-ElDP	468	RCN53435	30.99	EGF-like domain protein[*Ancylostoma caninum*]
Rs-VpSA	936	KHN71005	56.41	Vacuolar protein sorting-associated protein 45 [*Toxocara canis*]
Rs-GH	774	XP_024506972	35.07	Glycoside hydrolase, superfamily domain-containing protein [*Strongyloides ratti*]
Rs-PL	690	BAE48370	37.74	pectate lyase [*Bursaphelenchus xylophilus*]
Rs-ReCa2	836	KHN72759	57.50	Reticulocalbin-2 [*Toxocara canis*]

Table 19: NCBI accession number, amplicon size and description of the sequenced target genes of *R. similis*

Target Gene Id	Accession No.	Length	Description
Rs-AB	MN184224	1104	abhydrolase domain-containing protein
Rs-Ac16	MN184227	249	immunodominant hypodermal antigen Ac16
Rs-AsLM	MN184226	996	Astacin-like metalloendopeptidase
Rs-CDh	MN184225	354	carbonate dehydratase, eukaryotic-type
Rs-ElDP	MN184223	468	EGF-like domain protein
Rs-VpSA	MN184222	936	Vacuolar protein sorting-associated protein 45
Rs-GH	MN184221	774	Glycoside hydrolase, superfamily domain-containing protein
Rs-PL	MN184220	690	pectate lyase
Rs-ReCa2	MN184219	836	Reticulocalbin-2

3.2. Protein sequence analysis

Prediction of putative protein sequences for all target genes by translating the cDNA sequences using the EMBOSS program TRANSEQ resulted in polypeptides of varying lengths ranging from 82 amino acids to 367 amino acid residues. The amino acid composition of Rs-AB protein showed the highest number of alanine residues (29) whereas tryptophan (3) had the leastnumber of residues. The amino acid composition of putative protein ofRs-Ac16showed the presence of highest number of glycine residues (8) while methionine and tyrosine residues were completely absent. The amino acid composition of Rs-AsLM protein showed the highest number of argininеresidues (25) whereas tryptophan (4) had the least number of residues. The amino acid composition of putative protein of Rs-CDh showed the presence of highest number of glycine residues (13) while threonine (2) and tryptophan (2) residues were present in least numbers. The amino acid composition of Rs-ElDP protein showed the highest number of glycineresidues (19) whereas only one tryptophan residue (1) was observed. The amino acid composition of putative protein of Rs-VpSA showed the presence of highest number of leucine residues (43) whereas only one tryptophan residue (1) was observed. The amino acid composition of Rs-GH protein showed the highest number of serineresidues (26) whereas methionine (3) had the least number of residues. The amino acid composition of putative protein of Rs-PL showed the presence of highest number of glycine and serine residues (25) whereas only one methionine residue (1) was observed. The amino acid composition of putative protein of Rs-ReCa2 showed the presence of highest number of glutamic acid residues (29) while cysteine residues were completely absent.

Table 20: Predicted properties of target proteins in *R.similis* using the EMBOSS program TRANSEQ

Protein id	No. of AAs	Molecular weight (KDa)	Isoelectric point (pI)	Negative residues	Positive residues	Instability index	Aliphatic Index	GRAVY	Half-life
Rs-AB	367	40622.06	6.28	38	35	40.77	78.66	-0.202	30 min (yeast, in vivo)
Rs-Ac16	82	9281.40	8.94	12	14	48.69	72.44	-0.821	>20 hours (yeast, in vivo)
Rs-AsLM	331	38030.14	9.10	32	41	51.28	65.32	-0.794	>20 hours (yeast, in vivo)

Rs-CDh	118	13363.92	5.88	15	9	45.11	83.39	-0.452	10 min (yeast, in vivo)
Rs-ElDP	155	16763.94	8.04	11	13	47.21	57.94	-0.339	2 min (yeast, in vivo)
Rs-VpSA	312	35448.58	8.64	34	37	41.96	97.50	-0.342	>20 hours (yeast, in vivo)
Rs-GH	257	29020.01	6.40	25	24	50.13	76.34	-0.300	>20 hours (yeast, in vivo)
Rs-PL	229	24368.22	8.25	16	18	33.01	75.76	-0.269	>20 hours (yeast, in vivo)
Rs-ReCa2	278	32305.26	5.56	55	42	50.83	70.58	-0.861	2 min (yeast, in vivo)

The instability indices of proteins were found to be within the range of stability as all values were found to be more or less close to 40, and any protein whose instability index is smaller than 40 is predicted as stable. Values observed higher than 40 may be mostly due to presence of partial protein sequences. The highest instability index was observed for Rs-AsLM which was 51.28. The half-life is a prediction of the time taken for half of the amount of protein in a cell to disappear after its synthesis in the cell. For five target genes, Rs-Ac16, Rs-AsLM, Rs-VpSA,Rs-GH and Rs-PL the half-life was predicted to be >20 hours (yeast, *in vivo*) whereas Rs-AB had an estimated half-life of 30 min (yeast, *in vivo*) and Rs-CDh had 10 min (yeast, *in vivo*). The least half-life estimated was for Rs-VpSA and Rs-ReCa2 which was 2 min (yeast, *in vivo*). The results obtained from Expasy-Protparam tool is shown in Table 20. Analysis by NCBI Conserved Domain Database for the protein Cox1 revealed the presence of conserved multi domain Cytochrome C and Quinol oxidase polypeptide I and similarity with the superfamily Heme-copper oxidase subunit I. The putative protein sequence of ND1 gene showed the presence of NADH super family (Table 21).

Table 21: Conserved domains of target proteins of *R. similis*predicted using the EMBOSS program TRANSEQ

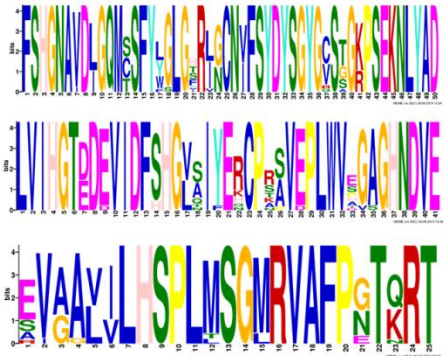
Protein id	Domain Name	Domain accession	Description	Interval	E-value
Rs-AB	MhpC	COG0596	Pimeloyl-ACP methyl ester carboxylesterase [Coenzyme transport and metabolism, General function prediction only]	109-420	7.65e-04
	PHA02857	PHA02857	monoglyceride lipase; Provisional	355-729	7.41e-03
	FrsA	COG1073	Fermentation-respiration switch protein FrsA, has esterase activity, DUF1100 family [Signal transduction mechanisms]	100-681	2.47e-21
Rs-Ac16	bacter_Pnkp super family	cl37383	polynucleotide kinase-phosphatase	46-162	4.34e-03
Rs-AsLM	ZnMc_astacin_like	cd04280	Zinc-dependent metalloprotease, astacin_like subfamily or peptidase family M12A, a group of zinc-dependent proteolytic enzymes with a HExxH zinc-binding site/active site	256-843	1.03e-37
	Astacin	pfam01400	Astacin (Peptidase family M12A)	256-849	1.28e-34
	ZnMc	smart00235	Zinc-dependent metalloprotease; Neutral zinc metallopeptidases.	247-702	2.62e-13

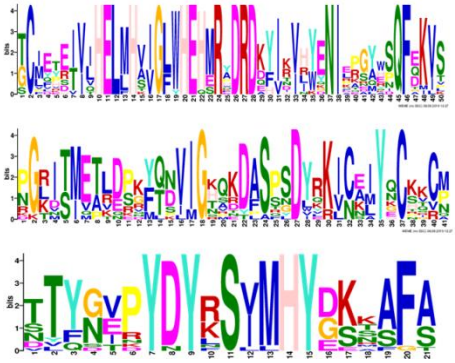
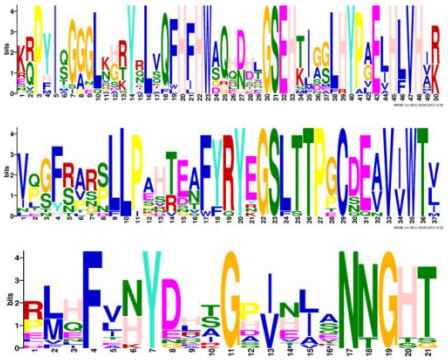
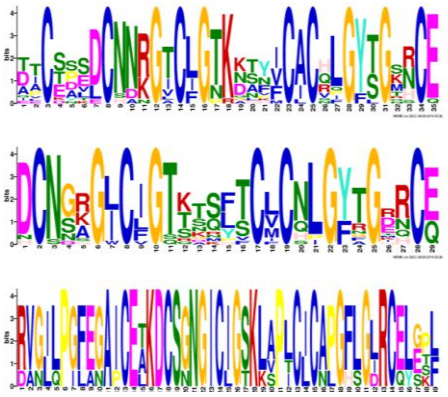
Rs-CDh	Carb_anhydrase	pfam00194	Eukaryotic-type carbonic anhydrase	10-351	1.87e-20
	alpha_CA	cd00326	Carbonic anhydrase alpha (vertebrate-like) group	70-351	3.54e-20
	Carb_anhydrase	smart01057	Eukaryotic-type carbonic anhydrase	67-351	6.96e-20
	Cah	COG3338	Carbonic anhydrase [Inorganic ion transport and metabolism]	67-297	1.01e-05
Rs-VpSA	SEC1	COG5158	Proteins involved in synaptic transmission and general secretion, Sec1 family	19-783	2.42e-40
	Sec1	pfam00995	Sec1 family	37-783	1.07e-25
Rs-GH	GH25_Lys1-like	cd06416	Lys-1 is a lysozyme encoded by the Caenorhabditis elegans lys-1 gene.	142-729	4.18e-20
	GH25_muramidase super family	cl10448	Endo-N-acetylmuramidases (muramidases) are lysozymes (also referred to as peptidoglycan hydrolases)	142-717	4.15e-32
Rs-PL	Pectate_lyase	pfam03211	Pectate lyase	130-651	2.60e-32
	Pectate_lyase super family	cl03937	Pectate lyase	130-651	1.39e-42
Rs-ReCa2	EFh_CREC_RCN2_like	cd16227	EF-hand, calcium binding motif, found in reticulocalbin-2 (RCN2) mainly from protostomes	6-542	1.54e-64
	EF-hand_7	pfam13499	EF-hand domain pair;	264-449	3.33e-06
	FRQ1	COG5126	Ca2+-binding protein, EF-hand superfamily [Signal transduction mechanisms]	12-458	6.50e-05
	PTZ00184	PTZ00184	calmodulin; Provisional	123-458	2.18e-03
	EFh_CREC super family	cl25354	EF-hand, calcium binding motif, found in CREC-EF hand family	6-545	3.98e-72

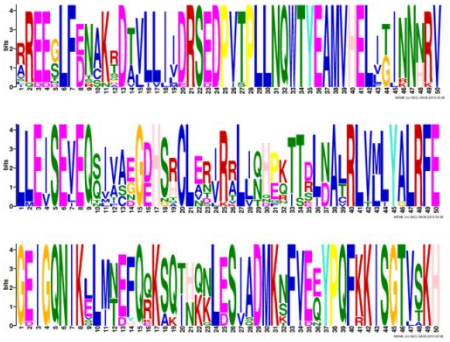
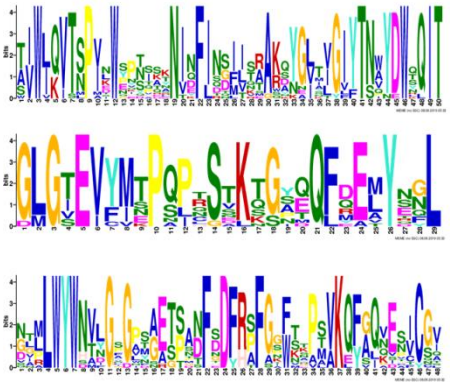
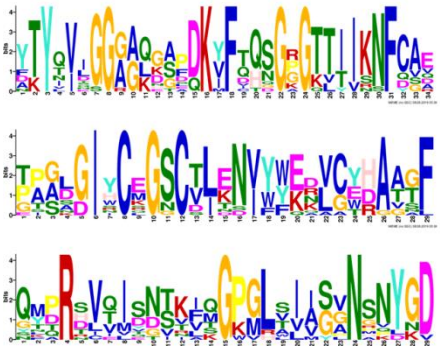
3.3. Motif characterization

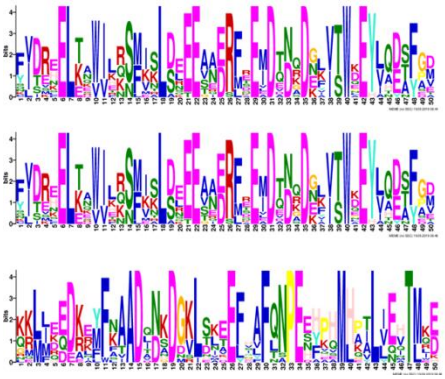
Sequence alignment of each target proteins with corresponding known nematode proteins wascarried out and was utilized for motif characterization. Structural motifs have been characterized with motif elucidation and sequence analysis methods. The structural motif characterization of target genes with corresponding nematode genes indicated the diversity of amino acids in the motifs of each targets. The comparison of structural features for all the protein sequences revealed that the proteins retained most of the conserved motifs that are crucial in their functions. Various motifs were identified to be conserved for all the target proteins. Conserved motif sequence logos of all the proteins predicted using MEME is shown in Table 22.

Table 22: Sequence logo of prominent motifs present in corresponding target proteins of *Radopholus similis*

Rs-AB	**EJW78952 W.bancrofti, CDJ88804 H.contortus, XP_013304149 N.americanus, KJH43748 D.viviparus, OZC05585 O.flexuosa, RCN37647 A.caninum, XP_003373594 T.spiralis, CDW57459 T.trichiura, KDR13312 Z.nevadensis, XP_026738972 T.ni, XP_012259616 A.rosae, XP_018320636 A.planipennis, XP_013197083 A.transitella,**	 **Rs-AB Motif 1** E-value- 1.16-650 **Rs-AB Motif 2** E-value- 1.2e-518 **Rs-AB Motif 3** E-value- 5.8e-288

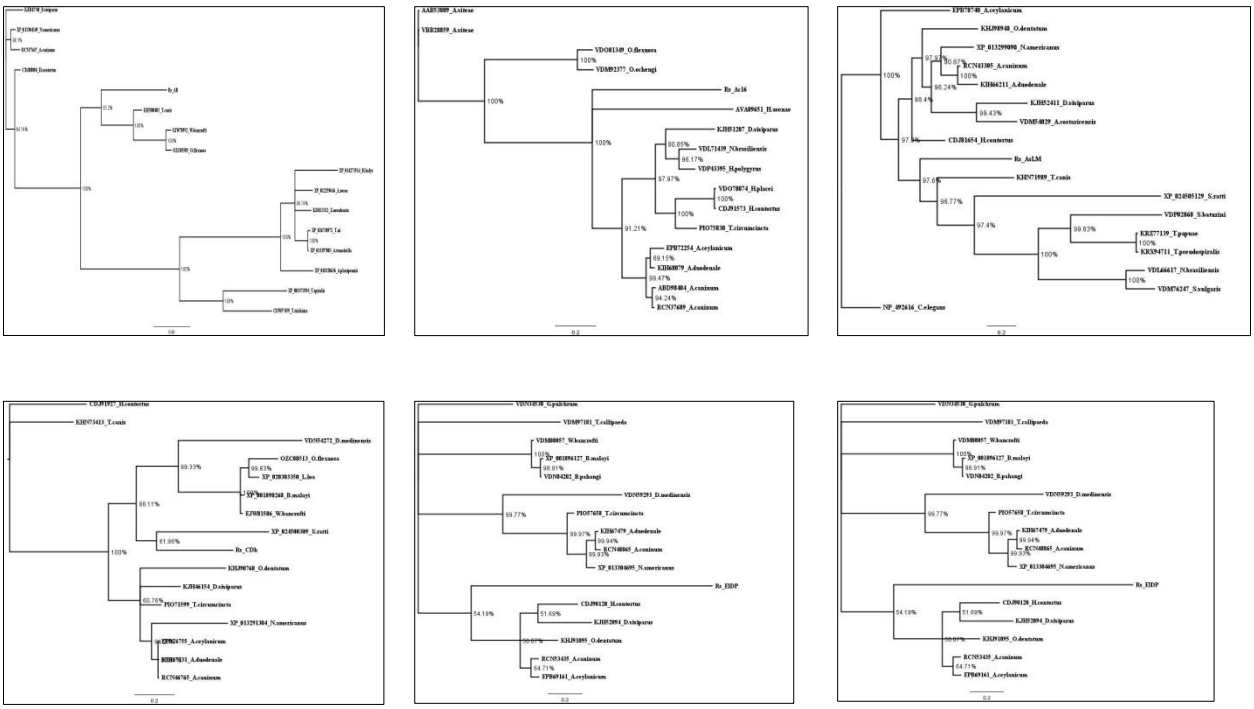
	XP_014273514 H.halys, KHN80403 T.canis		
Rs-AsLM	**KHJ98948 O.dentatum, XP_013299090 N.americanus, RCN43305 A.caninum, KJH52411 D.viviparus, KIH66211 A.duodenale, EPB78740 A.ceylanicum, XP_024505129 S.ratti, KHN71989 T.canis, NP_492616 C.elegans, KRZ77139 T.papuae, KRX94711 T.pseudospiralis, CDJ81654 H.contortus, VDM54029 A.costaricensis, VDL66617 N.brasiliensis, VDM76247 S.vulgaris, VDP02868 S.baturini**		**Rs-AsLM Motif 1** E-value- 3.7e-501 **Rs-AsLM Motif 2** E-value- 4.8e-248 **Rs-AsLM Motif 3** E-value- 1.1e-193
Rs-CDh	**PIO71599 T.circumcincta, XP_001898268 B.malayi, KJH46154 D.viviparus, EPB76755 A.ceylanicum, KIH67331 A.duodenale, XP_024500309 S.ratti, RCN46765 A.caninum, XP_020303350 L.loa, OZC08513 O.flexuosa, CDJ91927 H.contortus, EJW81586 W.bancrofti, KHN73413 T.canis, KHJ90768 O.dentatum, VDN54272 D.medinensis, XP_013291304 N.americanus**		**Rs-CDh Motif 1** E-value-1.5e-439 **Rs-CDh Motif 2** E-value-2.8e-309 **Rs-CDh Motif 3** E-value-4.4e-110
Rs-ElDP	**VDM97181 T.callipaeda, VDN59293 D.medinensis, VDN34530 G.pulchrum, RCN53435 A.caninum, CDJ90120 H.contortus, KHJ91095 O.dentatum, EPB69161 A.ceylanicum, KJH52094 D.viviparus, XP_001896127 B.malayi, KIH67479 A.duodenale, RCN48865 A.caninum, PIO57658 T.circumcincta, XP_013304695 N.americanus, VDM08057 W.bancrofti, VDN84202 B.pahangi**		**Rs-ElDP Motif 1** E-value-7.1e-226 **Rs-ElDP Motif 2** E-value-1.7e-193 **Rs-ElDP Motif 3** E-value-3.6e-065

Rs-VpSA	**VDO28940 O.flexuosa, VDK58161 A.simplex, XP_020305667 L.loa, CDJ91295 H.contortus, KHJ92134 O.dentatum, EPB72407 A.ceylanicum, VDL77850 N.brasiliensis, VDO64401 H.polygyrus, VDO83913 H.placei, VDM60135 A.costaricensis, RCN23762 A.caninum, XP_024508257 S.ratti, KHN71005 T.canis, XP_001898005 B.malayi, EJW85771 W.bancrofti**		**Rs-VpSA Motif 1** E-value-7.3e-584 **Rs-VpSA Motif 2** E-value-1.3e-566 **Rs-VpSA Motif 3** E-value-3.5e-461
Rs-GH	**XP_024506972 S.ratti, CDJ84788 H.contortus, NP_505642 C.elegans, XP_013308228 N.americanus, KHJ89924 O.dentatum, EYC08345 A.ceylanicum, KIH55997 A.duodenale, RCN46694 A.caninum, AEZ67029 S.papillosus, XP_003139373 L.loa, KJH51387 D.viviparus, VDL82009 N.brasiliensis, VDM94115 O.ochengi, XP_013299425 N.americanus, VDM81290 S.vulgaris, VDD88146 E.vermicularis**		**Rs-GH Motif 1** E-value-3.2e-365 **Rs-GH Motif 2** E-value-2.8e-179 **Rs-GH Motif 3** E-value-1.6e-370
Rs-PL	**BAE48370 B.xylophilus, BAE48375 B.mucronatus, AAQ97032 M.incognita, BAI44499 A.avenae, AFL48198 D.destructor, ADD82848 H.avenae, AIC75353 M.graminicola, AEA08828 G.virginiae, AEA08855 G.pallida, AAF80746 G.rostochiensis, AEA08814 G.tabacum tabacum, ALB38961 M.enterolobii, AAW56829 M.incognita, AAM74954 H.glycines**		**Rs-PL Motif 1** E-value-1.5e-200 **Rs-PL Motif 2** E-value-1.7e-191 **Rs-PL Motif 3** E-value-3.8e-117

Rs-ReCa2	**KHN72759 T.canis, XP_024504244 S.ratti, KJH48586 D.viviparus, XP_001901161 B.malayi, RCN48571 A.caninum, OZC12551_ O.flexuosa, XP_013302197 N.americanus, EJW87586 W.bancrofti, XP_028156606 O.furnacalis, XP_002067188 D.willistoni, XP_020806068 D.serrata, XP_017063241 D.eugracilis, XP_017043794 D.ficusphila, XP_021197120 H.armigera, XP_015794895 T.urticae, CAD59175 C.elegans**	 **Rs-ReCa2 Motif 1** E-value-6.3e-305 **Rs-ReCa2 Motif 2** E-value-3.3e-305 **Rs-ReCa2 Motif 3** E-value-7.4e-308

3.4. Evolutionary analysis

Phylogenetic relationships within and between various parasitic nematode species were carried out on each target sequences by means of Bayesian Inference (BI), Maximum Likelihood (ML) and Maximum Parsimony (MP) methods without using any outgroups or using *C. elegans* as outgroup wherever available. All the three trees obtained were used to generate the consensus-based phylogeny as shown in Figure 20.



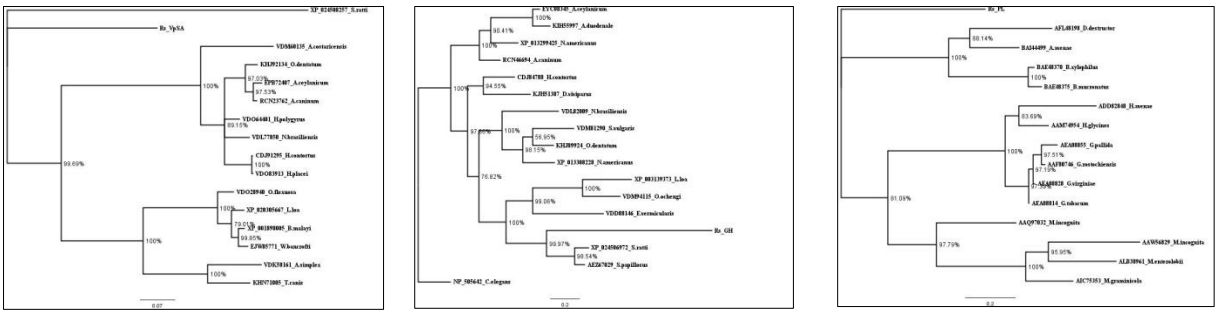


Figure 20: Consensus phylogenetic tree of amino-acid sequence for inferring relationship within the genes Rs-AB, Rs-Ac16, Rs-AsLM, Rs-CDh, Rs-ElDP, Rs-VpSA, Rs-GH, Rs-PL and Rs-ReCa2 against other parasitic nematodes obtained by Bayesian analysis, Maximum likelihood and Maximum parsimony methods

4. EXPERIMENTAL VALIDATION OF NOVEL TARGET GENES

4.1. Assessment of dsRNA treatment

Significant differences were seen in the dsRNA-soaked nematodes, many of which lost the typical body shape and displayed abnormal locomotor behavior when compared to untreated nematodes. The nematodes that showed no movements when touched were considered dead. After incubation for 4 h in dsRNA solution with FITC (as a visible marker), strong fluorescence signals were observed in the stylet, esophageal lumen, intestine, amphids, excretory duct and anus of *R. similis* (Figure 21). Fluorescence microscopy also indicated that the soaking solution diffused into the body of *R. similis* and did spread more evenly when observed after 24h. After 44 h of incubation nematodes were mostly dead and lost the typical body shape in case of all genes.

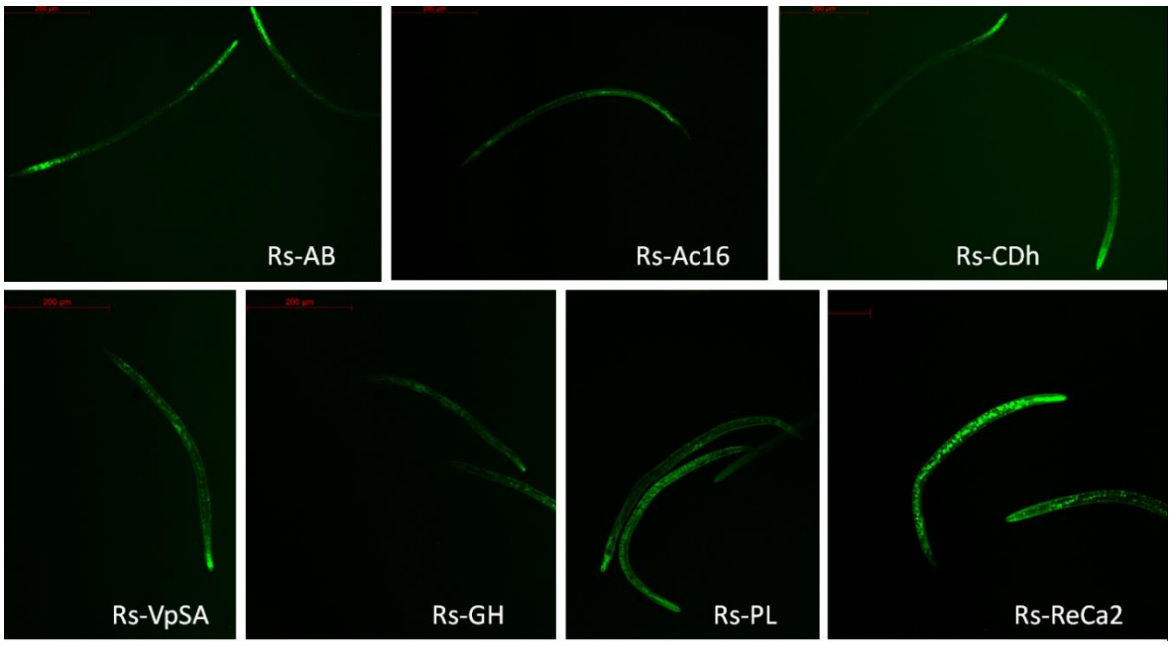


Figure 21: FITC fluorescence and viability in mixed-stages of *R. similis* incubated in soaking solution containing FITC and dsRS-AB after 4hours

4.2. Gene expression analysis

Changes in the expressions of the seven genes in 4h, 24h and 48 h after RNAi treatment on comparing with those in untreated nematodes, were not consistent as observed by Iqbal et al.,. (2020). Although RNAi treatment is expected to result in down-regulation of target genes, gene expression was significantly reduced only in 3 out of 7 genes silenced. There was ~32 and ~10-fold reduction in expression of Rs-GH gene after 24h and 48 h soaking and ~15 and ~8-fold reduction in expression was seen in in Rs-AB and Rs-PL genes after 24h treatment (Figure 22). No expression was observed in case of Rs-CDh and Rs-PL after 48h incubation.

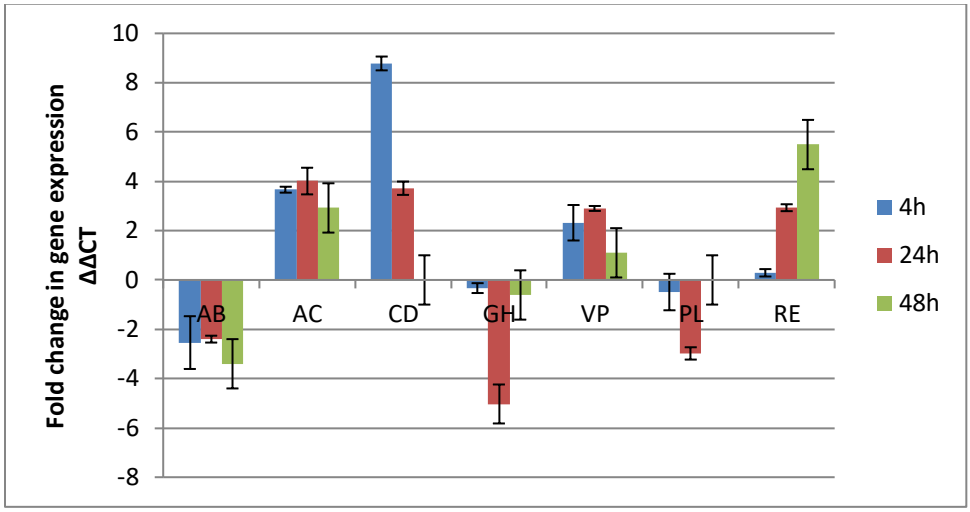


Figure 22: Fold change in target gene expression of *R.similis* 4h, 24h and 48 h after soaking in dsRNA of target genes

4.3. Infectivity and reproductive bioassay

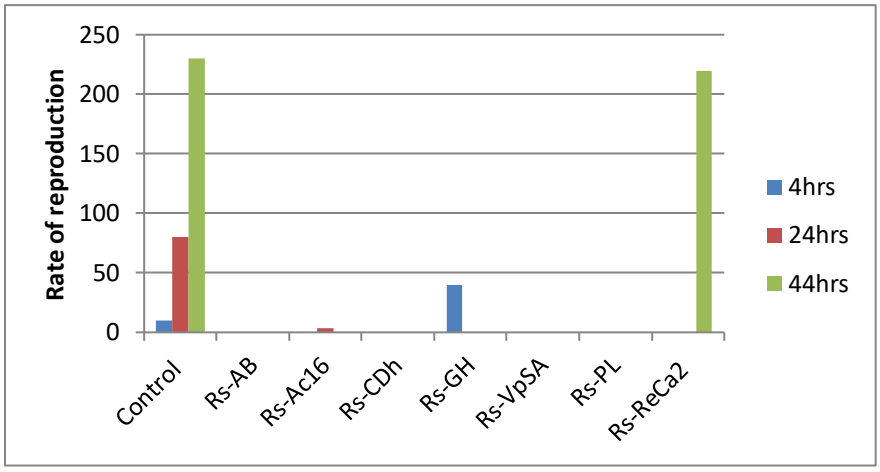


Figure 23: Reproductive rates of *R. similis* extracted from carrot disks at 50 days after inoculation with 30 females.

The effect of silencing on the reproductive capabilities of *R. similis* was assessed by inoculating the nematodes on carrot disks. After culturing on carrot disks for 50 d, nematodes treated with dsRNAs for 4, 24, 48 h had a significantly lower reproductive rate (reproductive rate = final nematodes⁄ initial nematodes) than untreated nematode (Figure 23).

5. DEVELOPMENT OF BIOINFORMATICS PIPELINE

An important aspect of dealing with this vast amount of data generated using technologies like RNA-seq is the processing method used to extract and intercept the information and most of these methods contains different steps, some of which are computationally intensive. Although Bioconductor offers a large number of packages for the statistical analysis of RNA-seq data, but most of them focuses on one or two specific aspects of data analysis. Combining them appropriately requires a good understanding of their functionalities and depending on the data, different combinations of these packages have to be used leading to a tough learning curve for aRNA-seq novice.

Here we introduce an integrated R based RNA-seq analysis pipeline, RNA-seqAnalysis, providing utilities for running an automated end-to-end analysis workflow to ease out the processing of RNA-seq data, hiding the intricate interplay of different packages behind simple functionalities that can be easily implemented just with the call of simple R functions. In this workflow, RNA-sequencing data from the *C. elegans*, was analyzed using popular biocounductor package to import, filter, normalize and systematize the data, followed by differential gene expression, go term annotation and splice junction analysis as a single workflow using eight different R functions. This pipeline further offers interactive exploration and easy storage of the results in form of plots and spreadsheets so that individual samples and genes can be later on easily examined by the user (Figure 24).

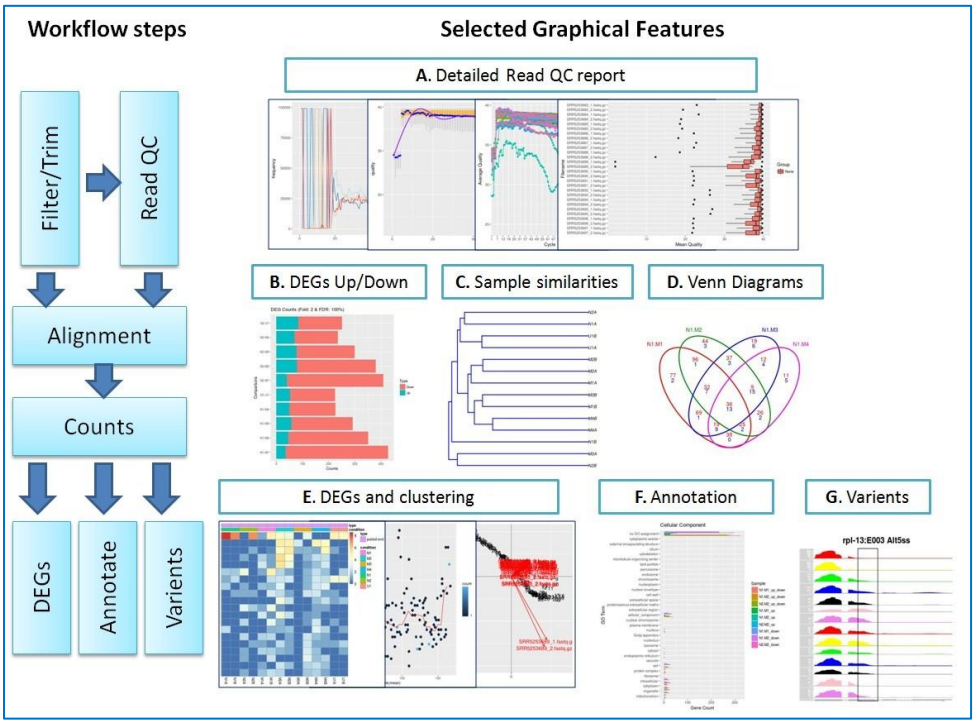


Figure24:Relevant workflow steps involved RNA-seq analysis and some selected graphical results generated through the pipeline

The pipeline is uploaded in Git-hub as a bioconductor package for review and is freely available for download (Figure 25). The working has been demonstrated and explained in the form of vignettes (user manual).

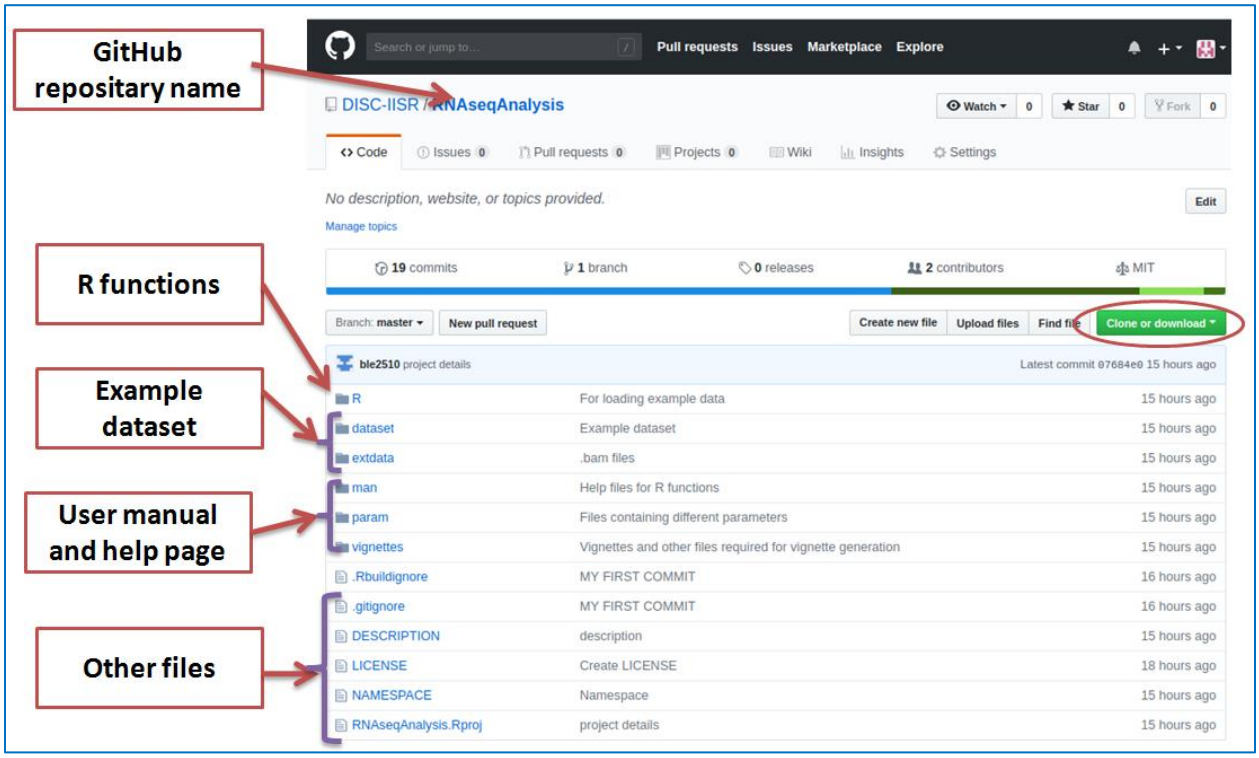


Figure 25: GitHub homepage showing RNA-seq analysis pipeline

CONCLUSION

Molecular characterization of *R. similis*: Considering the differences in rates of evolution of different marker genes and the low resolution of ITS1 gene, a larger number of genes were used in the present study for the phylogenetic analysis instead of a single marker gene.Two mitochondrial genes (cox1 and NADH) along with internal transcribed spacer region (ITS1-5.8S-ITS2), D3 expansion domain segments, 18S ribosomal RNA gene and 28S ribosomal RNA gene were used for the molecular characterization of the Indian isolate of *R. similis*. Based on the BLAST analysis and MLST studies the identity of the Indian isolate was confirmed as *Radopholus similis*. The identified proteins from the mitochondrial genes can act as a valuable resource towards development of more stringent nematode identification methods.The present study is the first to clone and characterize the mitochondrial genes- cytochrome c oxidase subunit 1 and NADH dehydrogenase subunit 1 from *R. similis*. The study has clearly proved the molecular identity of black pepper isolate of burrowing nematode from India as *R. similis*. However, the highest percentage of polymorphic sites and divergence was observed in ITS sequence indicating high degree of selection pressure. However, similar analysis could not be extended to mitochondrial genes for want of sufficient data from other populations.

Transcriptome sequencing and ES protein identification: In this study, we have sequenced and annotated the transcriptome of mixed stages of *R. similis* by Illumina sequencing. The final transcriptome assembly contained ~60,000 transcripts having 97% similarity at the nucleotide level. This is much more than normally seen in other recently studied PPNs transcriptome, and may be largely due to the presence of several isoforms that could not be collapsed into single genes. As in other PPN species, our Blast analyses indicated that a significant number of candidate genes lack annotation and a predicted function. A total of only 11.8% proteins were classified as "enzymes" according to KEGG. Gene Ontology mapping revealed that most genes have basic functions such as transport, transcription, protein synthesis or modification and developmental and metabolic processes. Further work is needed to characterize them and to identify their possible functional relevance. To be a successful plant pathogen, nematodes must be able to successfully locate a host plant root, avoid induction of an effective host immune response, resist any immune response that is expressed and able to manipulate host's developmental and cell biology to enter the plant cell. These interactions are mediated by an armory of effectors or the ES proteins that plant-parasitic nematodes appear to have assembled from adaptation of endogenous genes and also loci acquired through horizontal gene transfer from a diverse range of other taxa (Eves-van *et al.,* 2016). To predict new or putative secreted proteins encoded by *R. similis* contigs with signal peptides and carbohydrate enzymatic activity or binding modules but with no transmembrane helices/domain need to be identified. Several plant cellwall modifying proteins were identified in *R. similis,* which are essential for their migration and feeding. Our dataset extends the arsenal of enzymes in *R. similis* with pectate lyase, alpha-sialyltransferase, xyloglucan, chitinase, polygalacturonase and arabinogalactan galactosidase.

Identification and sequencing of *R. similis* target genes: From among the ES proteins, seven genes were randomly selected, amplified, sequenced and annotated for presence of conserved functional domains.Out of these, three genes that showed significant decrease in transcript abundance have important functional domains and conserved families. GHs are known to be present in a wide range of organisms including bacteria, plants, fungi and nematodes (Tyler *et al.,* 2010), but within PPNs, GHs are known to be involved in functions like breakdown of plant cell walls to aid the nematode to infect the host tissues and to ease their movement within the host cells. Additionally, GHs also aid in the breakdown of simple and complex carbohydrates such as glucose and sucrose, and thus provide potential resource for nutrition of nematodes. The putative GHs in *R. similis* belong to multiple families, notably, GH5 (cellulase), GH25 (lysozyme) and GH30 (endo-beta-1,4-xylanase). In this study, the sequenced Rs-GH showed the presence of two highly conserved domains GH25_Lys1-like and GH25_muramidase super family. This study is the first report of GH gene containing GH25 family in *R. similis*. Pectate lyases play a key role in the degradation of the most abundant components in plant cell wall, pectin (Caffall and Mohnen, 2009). Pectate lyases which play an important pathogenicity or virulence factor, are extensively distributed in different families of microorganisms (Payasi *et al.,* 2009). Recently, studies have suggested that PPN pectate lyases can promote nematode infection, as silencing of pectate lyase genes, Hs-PEL2, Hg-PEL6 and Me-PEL1 significantly reduced the ability of nematodes to infect host (Vanholme *et al.,* 2007; Zhuo *et al.,* 2011; Peng *et al.,* 2016). In the present study, Rs-PL that showed presence of conserved Pectate_lyase super family, showed decrease in transcript abundance. Abhydrolase domain-containing protein have not been studied much in plant parasitic nematodes

except in the genome of *Meloidogyne graminicola.* Rs-AB showed presence of conserved families like MhpC, PHA02857 andFrsA.This study provides the baseline for selection of future nematode gene candidates for host-induced gene silencing, eventually to design effective control strategies against this nematode having a broad host range.

Experimental validation by RNAi approach: Transcript abundance of the seven genes following soaking in dsRNA for 4h, 24 h and 48h gave varied results; Rs-AB, Rs-GH and Rs-PL showed significant decreases whereas other four genes (Rs-Ac16, Rs-CDh, Rs-VpSA and Rs-ReCa2) showed significant increases. In similar RNAi studies with other PPNs, a reduction in gene expression is reported (Tan *et al.,* 2013; Li *et al.,* 2015; Fosu-Nyarko *et al.,* 2016; Wang *et al.,* 2016; Blyuss *et al.,* 2019; Tian *et al.,* 2019). However, as in the present study, gene silencing was not observed in a few studiesinvolving*Meloidogyne*, *Heterodera*, and *Globodera* spp. indicating that some genes are refractory to exogenous RNAi with either no measurable reduction or increase in gene expression (Bakhetia *et al.,* 2008; Dalzell *et al.,* 2010; Danchin *et al.,* 2013; Chi *et al.,* 2016; Shivakumara *et al.,* 2016; Iqbal *et al.,* 2020). In such cases their function may not essentially be the cause of the apparent resistance to exogenous RNAi. Alternatively, some aspects like the quantity of dsRNA used as triggers for RNAi, the presence of paralogs genes, differential rate of target gene expression, the nematodes' susceptibility to RNAi or the need to sustain homeostasis subsequent to an initial change in gene expression may have a role in the ability of nematode to resist RNAi or to recover from its effect (Iqbal *et al.,*2020). The recalcitrance to exogenous RNAi or recovery from the effects of RNAi in PPNs due to the above-mentioned factors has not yet been extensively studied.On the other hand, for those genes where silencing considerably affected the biology of the nematodes, it can be concluded that the amount of the dsRNA used efficiently initiated RNAi. For the studied genes where silencing triggered some considerable RNAi effect on the nematode, there is the potential to use these genes as targets for nematode control using different RNAi approaches.

REFERENCES

1. Alkharouf, N. W., Klink, V. P. and Matthews, B. F. (2007). Identification of *Heterodera glycines* (soybean cyst nematode [SCN]) cDNA sequences with high identity to those of *Caenorhabditis elegans* having lethal mutant or RNAi phenotypes. *Experimental parasitology*, *115*(3), 247-258.
2. Altschul, S. F., Gish, W., Miller, W. C., Myers, E. W. and Lipman, D. J. (1990). Basic Local Alignment Search Tool. journal of molecular biology 215, 403–410, doi: 10.1016. *S0022-2836 (05)*, 80360-2.
3. Antonino de Souza Júnior, J. D., Ramos Coelho, R., Tristan Lourenço, I., da Rocha Fragoso, R., Barbosa Viana, A. A., Lima Pepino de Macedo, L., ... and Grossi-de-Sa, M. F. (2013). Knocking-down *Meloidogyne incognita* proteases by plant-delivered dsRNA has negative pleiotropic effect on nematode vigor. PLoS One, 8(12), e85364.
4. Arguel, M. J., Campan-Fournier, A., Perfus-Barbeoch, L., Magliano, M., Rosso, M. N., Da Rocha, M., ... and Abad, P. (2013). Identification of novel target genes for safer and more specific control of root-knot nematodes from a pan-genome mining. *PLOS Pathogens 10 (9), 1-15*.
5. Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., ... and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology*, *37*(4), 420-423.
6. Atkinson, H. J., and Harris, P. D. (1989). Changes in nematode antigens recognized by monoclonal antibodies during early infections of soya beans with the cyst nematode Heterodera glycines. *Parasitology*, *98*(3), 479-487.
7. Backman, T. W., and Girke, T. (2016). systemPipeR: NGS workflow and report generation environment. *BMC bioinformatics*, *17*(1), 1-8.
8. Bai, M., Qing, X., Qiao, K., Ning, X., Xiao, S., Cheng, X. and Liu, G. (2020). Mitochondrial COI gene is valid to delimitate Tylenchidae (Nematoda: Tylenchomorpha) species. *Journal of nematology*, 52, e2020-38.

9. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... and Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, *37*(suppl_2), W202-W208.
10. Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Research*, *43*(W1), W39-W49. Retrieved from: http://meme-suite.org/tools/meme.
11. Bakhetia, M., Urwin, P. E. and Atkinson, H. J. (2007). qPCR analysis and RNAi define pharyngeal gland cell-expressed genes of *Heterodera glycines* required for initial interactions with the host. *Molecular Plant-Microbe Interactions*, *20*(3), 306-312.
12. Bakhetia, M., Urwin, P. E. and Atkinson, H. J. (2008). Characterisation by RNAi of pioneer genes expressed in the dorsal pharyngeal gland cell of *Heterodera glycines* and the effects of combinatorial RNAi. *International journal for parasitology*, *38*(13), 1589-1597.
13. Bauters, L., Haegeman, A., Kyndt, T. and Gheysen, G. (2014). Analysis of the transcriptome of *Hirschmanniella oryzae* to explore potential survival strategies and host–nematode interactions. *Molecular plant pathology*, *15*(4), 352-363.
14. Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics*, *33*(16), 2583-2585.
15. Bellafiore, S., Shen, Z., Rosso, M. N., Abad, P., Shih, P. and Briggs, S. P. (2008). Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. *PLoS pathogens*, *4*(10), e1000192.
16. Blouin, M. S. (2002). Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer. *International journal for parasitology*, *32*(5), 527-531.
17. Blyuss, K. B., Fatehi, F., Tsygankova, V. A., Biliavska, L. O., Iutynska, G. O., Yemets, A. I. and Blume, Y. B. (2019). RNAi-based biocontrol of wheat nematodes using natural poly-component biostimulants. *Frontiers in plant science*, *10*, 483.
18. Bolla, R. I., Weaver, C. and Winter, R. E. K. (1988). Genomic differences among pathotypes of Bursaphelenchusxylophilus. *Journal of Nematology*, *20*(2), 309.
19. Bonin-Debs, A. L., Boche, I., Gille, H., and Brinkmann, U. (2004). Development of secreted proteins as biotherapeutic agents. *Expert opinion on biological therapy*, *4*(4), 551-558.
20. Buffalo, V. (2012). QRQC-quick read quality control. *Reference Source*.
21. Caffall, K. H. and Mohnen, D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate research*, *344*(14), 1879-1900.
22. Channale, S., Kalavikatte, D., Thompson, J. P., Kudapa, H., Bajaj, P., Varshney, R. K., ... and Thudi, M. (2021). Transcriptome analysis reveals key genes associated with root-lesion nematode *Pratylenchus thornei* resistance in chickpea. *Scientific Reports*, *11*(1), 1-11.
23. Charlton, W. L., Harel, H. Y. M., Bakhetia, M., Hibbard, J. K., Atkinson, H. J. and McPherson, M. J. (2010). Additive effects of plant expressed double-stranded RNAs on root-knot nematode development. *International Journal for Parasitology*, 40(7), 855-864.
24. Chen, Q., Rehman, S., Smant, G. and Jones, J. T. (2005). Functional analysis of pathogenicity proteins of the potato cyst nematode *Globodera rostochiensis* using RNAi. *Molecular Plant-Microbe Interactions*, *18*(7), 621-625.
25. Chi, Y., Wang, X., Le, X., Ju, Y., Guan, T., and Li, H. (2016). Exposure to double-stranded RNA mediated by tobacco rattle virus leads to transcription up-regulation of effector gene Mi-vap-2 from Meloidogyne incognita and promotion of pathogenicity in progeny. *International journal for parasitology*, *46*(2), 105-113.
26. Cho, S. J., Jeftinija, K., Glavaski, A., Jeftinija, S., Jena, B. P., and Anderson, L. L. (2002). Structure and dynamics of the fusion pores in live GH-secreting cells revealed using atomic force microscopy. *Endocrinology*, *143*(3), 1144-1148.
27. Cho, W. J., Jeremic, A., Rognlien, K. T., Zhvania, M. G., Lazrishvili, I., Tamar, B., and Jena, B. P. (2004). Structure, isolation, composition and reconstitution of the neuronal fusion pore. *Cell biology international*, *28*(10), 699-708.
28. Choudhary, D., Koulagi, R., Rohatagi, D., Kumar, A., Jain, P. K. and Sirohi, A. (2012). Engineering resistance against root-knot nematode, *Meloidogyne incognita*, by host delivered RNAi. In International Conference on Plant Biotechnology for food Security: New Frontiers (New Delhi: National Agricultural Science Centre) (pp. 21-24).
29. Clement, M., Snell, Q., Walker, P., Posada, D. and Crandall, K. (2002). TCS: estimating gene genealogies. In *Parallel and Distributed Processing Symposium, International* (Vol. 3, pp. 0184-0184). IEEE Computer Society.
30. Cotton, J. A., Lilley, C. J., Jones, L. M., Kikuchi, T., Reid, A. J., Thorpe, P., ... and Urwin, P. E. (2014). The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome biology*, *15*(3), 1-17.
31. Curtis, R. H. (2007). Plant parasitic nematode proteins and the host–parasite interaction. *Briefings in Functional Genomics and Proteomics*, *6*(1), 50-58.
32. Da Rocha, M., Bournaud, C., Dazenière, J., Thorpe, P., Bailly-Bechet, M., Pellegrin, C., ... and Danchin, E. G. (2021). Genome expression dynamics reveal the parasitism regulatory landscape of the root-knot nematode *Meloidogyne incognita* and a promoter motif associated with effector genes. *Genes*, *12*(5), 771.

33. Dalzell, J. J., McMaster, S., Fleming, C. C. and Maule, A. G. (2010). Short interfering RNA-mediated gene silencing in *Globodera pallida* and *Meloidogyne incognita* infective stage juveniles. *International journal for parasitology*, *40*(1), 91-100.
34. Danchin, E. G., Arguel, M. J., Campan-Fournier, A., Perfus-Barbeoch, L., Magliano, M., Rosso, M. N., ... and Guy, J. (2013). Identification of novel target genes for safer and more specific control of root-knot nematodes from a pan-genome mining. *PLOS Pathogens, 9*(10), e1003745
35. Darriba, D., Taboada, G. L., Doallo, R. and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, *9*(8), 772-772.
36. David, T. K., and Eric, L. D., (1990). Improved nematode extraction from carrot disk culture. *Journal of Nematology*, *22*(3), 399.
37. Davis, E. L., Hussey, R. S., Baum, T. J., Bakker, J., Schots, A., Rosso, M. N. and Abad, P. (2000). Nematode parasitism genes. *Annual review of phytopathology*, *38*(1), 365-396.
38. Delhomme, N., Padioleau, I., Furlong, E. E., and Steinmetz, L. M. (2012). easyRNA-seq: a bioconductor package for processing RNA-seq data. *Bioinformatics*, *28*(19), 2532-2533.
39. Dinh, P. T., Brown, C. R. and Elling, A. A. (2014a). RNA interference of effector gene Mc16D10L confers resistance against *Meloidogyne chitwoodi* in Arabidopsis and potato. *Phytopathology*, *104*(10), 1098-1106.
40. Dinh, P. T., Zhang, L., Brown, C. R. and Elling, A. A. (2014b). Plant-mediated RNA interference of effector gene Mc16D10L confers resistance against *Meloidogyne chitwoodi* in diverse genetic backgrounds of potato and reduces pathogenicity of nematode offspring. *Nematology*, *16*(6), 669-682.
41. Drummond, A. J., Suchard, M. A., Xie, D. and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution*, *29*(8), 1969-1973.
42. Dubreuil, G., Magliano, M., Deleury, E., Abad, P. and Rosso, M. N. (2007). Transcriptome analysis of root-knot nematode functions induced in the early stages of parasitism. *New Phytologist*, *176*(2), 426-436.
43. Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, *21*(16), 3439-3440.
44. Dutta, T. K., Papolu, P. K., Banakar, P., Choudhary, D., Sirohi, A. and Rao, U. (2015). Tomato transgenic plants expressing hairpin construct of a nematode protease gene conferred enhanced resistance to root-knot nematodes. *Frontiers in microbiology*, 6, 260.
45. Elashry, A. M., Habash, S. S., Vijayapalani, P., Brocke-Ahmadinejad, N., Blümel, R., Seetharam, A., ... and Grundler, F. M. (2020). transcriptome and parasitome Analysis of Beet cyst nematode Heteroderaschachtii. *Scientific reports*, *10*(1), 1-12.
46. Elbadri, G. A., De Ley, P., Waeyenberge, L., Vierstraete, A., Moens, M. and Vanfleteren, J. (2002). Intraspecific variation in *Radopholus similis* isolates assessed with restriction fragment length polymorphism and DNA sequencing of the internal transcribed spacer region of the ribosomal RNA cistron. *International Journal for Parasitology*, *32*(2), 199-205.
47. Emanuelsson, O., Nielsen, H., Brunak, S. and Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of molecular biology*, *300*(4), 1005-1016.
48. European and Mediterranean Plant Protection Organization (EPPO) Diagnostics Report (2008) *Radopholus similis*. EPPO Bull. 38, 374– 378.
49. European and Mediterranean Plant Protection Organization(1997)Recommendations made by EPPO Council in 1996. *Bulletin OEPP/EPPO Bulletin*, *27*, 643-644.
50. Eves-van den Akker, S., Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G., Da Rocha, M., ... and Jones, J. T. (2016). The genome of the yellow potato cyst nematode, Globoderarostochiensis, reveals insights into the basis of parasitism and virulence. *Genome biology*, *17*(1), 1-23.
51. Eves-van den Akker, S., Lilley, C. J., Danchin, E. G., Rancurel, C., Cock, P. J., Urwin, P. E., and Jones, J. T. (2014). The transcriptome of *Nacobbus aberrans* reveals insights into the evolution of sedentary endoparasitism in plant-parasitic nematodes. *Genome Biology and Evolution*, *6*(9), 2181-2194.
52. Fairbairn, D. J., Cavallaro, A. S., Bernard, M., Mahalinga-Iyer, J., Graham, M. W. and Botella, J. R. (2007). Host-delivered RNAi: an effective strategy to silence genes in plant parasitic nematodes. *Planta*, 226(6), 1525-1533.
53. Fan, W., Wei, Z., Zhang, M., Ma, P., Liu, G., Zheng, J., ... and Zhang, P. (2015). Resistance to *Ditylenchus destructor* infection in sweet potato by the expression of small interfering RNAs targeting unc-15, a movement-related gene. *Phytopathology*, 105(11), 1458-1465.
54. Fanelli, E., Di Vito, M., Jones, J. T. and De Giorgi, C. (2005). Analysis of chitin synthase function in a plant parasitic nematode, *Meloidogyne artiellia*, using RNAi. *Gene*, 349, 87-95.
55. Felsenstein, J. (1993). *PHYLIP (phylogeny inference package), version 3.5 c*. Joseph Felsenstein..
56. Felsenstein, J. (2013) PHYLIP (phylogeny inference package) version 3.695. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, WA.

57. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, *391*(6669), 806-811.
58. Fosu-Nyarko, J., Tan, J. A. C., Gill, R., Agrez, V. G., Rao, U., and Jones, M. G. (2016). D e novo analysis of the transcriptome of P ratylenchus zeae to identify transcripts for proteins required for structural integrity, sensation, locomotion and parasitism. *Molecular Plant Pathology*, *17*(4), 532-552.
59. Gahoi, S., Singh, S. and Gautam, B. (2019). Genome-wide identification and comprehensive analysis of Excretory/Secretory proteins in nematodes provide potential drug targets for parasite control. *Genomics*, *111*(3), 297-309.
60. Garg, G., and Ranganathan, S. (2012). Helminth secretome database (HSD): a collection of helminth excretory/secretory proteins predicted from expressed sequence tags (ESTs). In *BMC genomics* (Vol. 13, No. 7, pp. 1-11). BioMed Central.
61. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook*, 571-607.
62. Gleason, C. A., Liu, Q. L. and Williamson, V. M. (2008). Silencing a candidate nematode effector gene corresponding to the tomato resistance gene Mi-1 leads to acquisition of virulence. *Molecular Plant-Microbe Interactions*, 21(5), 576-585.
63. Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., ... and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research*, *36*(10), 3420-3435.
64. Hadley, W. (2016). *Ggplot2: Elegrant graphics for data analysis*. Springer.
65. Haegeman, A., Bauters, L., Kyndt, T., Rahman, M. M. and Gheysen, G. (2013). Identification of candidate effector genes in the transcriptome of the rice root knot nematode *Meloidogyne graminicola*. *Molecular Plant Pathology*, *14*(4), 379-390.
66. Haegeman, A., Jacob, J., Vanholme, B., Kyndt, T. and Gheysen, G. (2008). A family of GHF5 endo-1, 4-beta-glucanases in the migratory plant-parasitic nematode *Radopholus similis*. *Plant Pathology*, *57*(3), 581-590.
67. Haegeman, A., Joseph, S. and Gheysen, G. (2011). Analysis of the transcriptome of the root lesion nematode *Pratylenchus coffeae* generated by 454 sequencing technology. *Molecular and biochemical parasitology*, *178*(1-2), 7-14.
68. Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic Acids Symp. Ser.* (Vol. 41, pp. 95-98).
69. Handoo, Z. A., Carta, L. K., Skantar, A. M., Subbotin, S. A. and Fraedrich, S. W. (2016). Molecular and morphological characterization of *Xiphinema chambersi* population from live oak in Jekyll Island, Georgia, with comments on morphometric variations. *Journal of Nematology*, *48*(1), 20.
70. Harcus, Y. M., Parkinson, J., Fernández, C., Daub, J., Selkirk, M. E., Blaxter, M. L., and Maizels, R. M. (2004). Signal sequence analysis of expressed sequence tags from the nematode Nippostrongylus brasiliensis and the evolution of secreted proteins in parasites. *Genome biology*, *5*(6), 1-15.
71. Hekman, J. P., Johnson, J. L. and Kukekova, A. V. (2015). Transcriptome analysis in domesticated species: challenges and strategies. *Bioinformatics and biology insights*, *9*, BBI-S29334.
72. Holterman, M., Karssen, G., Van Den Elsen, S., Van Megen, H., Bakker, J. and Helder, J. (2009). Small subunit rDNA-based phylogeny of the Tylenchida sheds light on relationships among some high-impact plant-parasitic nematodes and the evolution of plant feeding. *Phytopathology*, *99*(3), 227-235.
73. Huang, X. and Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome research*, *9*(9), 868-877.
74. Huang, X., Xu, C. L., Yang, S. H., Li, J. Y., Wang, H. L., Zhang, Z. X., ... and Xie, H. (2019). Life-stage specific transcriptomes of a migratory endoparasitic plant nematode, *Radopholus similis* elucidate a different parasitic and life strategy of plant parasitic nematodes. *Scientific reports*, *9*(1), 1-11.
75. Iberkleid, I., Vieira, P., de Almeida Engler, J., Firester, K., Spiegel, Y. and Horowitz, S. B. (2013). Fatty acid-and retinol-binding protein, Mj-FAR-1 induces tomato host susceptibility to root-knot nematodes. *PloS one*, 8(5), e64586.
76. Ibrahim, H. M., Alkharouf, N. W., Meyer, S. L., Aly, M. A., Abd El Kader, Y., Hussein, E. H. and Matthews, B. F. (2011). Post-transcriptional gene silencing of root-knot nematode in transformed soybean roots. *Experimental parasitology*, 127(1), 90-99.
77. Iqbal, S., Fosu-Nyarko, J. and Jones, M. G. (2016). Genomes of parasitic nematodes (*Meloidogyne hapla*, *Meloidogyne incognita*, *Ascaris suum* and *Brugia malayi*) have a reduced complement of small RNA interference pathway genes: knockdown can reduce host infectivity of *M. incognita*. *Functional* and *integrative genomics*, *16*(4), 441-457.
78. Iqbal, S., Fosu-Nyarko, J. and Jones, M. G. (2020). Attempt to silence genes of the RNAi pathways of the root-knot nematode, *Meloidogyne incognita* results in diverse responses including increase and no change in expression of some genes. *Frontiers in plant science*, *11*, 328.
79. Jacob, J. E., Vanholme, B., Van Leeuwen, T. and Gheysen, G. (2009). A unique genetic code change in the mitochondrial genome of the parasitic nematode *Radopholus similis*. *BMC research notes*, *2*(1), 1-10.

80. Jaouannet, M., Magliano, M., Arguel, M. J., Gourgues, M., Evangelisti, E., Abad, P. and Rosso, M. N. (2013). The root-knot nematode calreticulin Mi-CRT is a key effector in plant defense suppression. *Molecular Plant-Microbe Interactions*, *26*(1), 97-105.
81. Jing, L., Guo, D., Hu, W., and Niu, X. (2017). The prediction of a pathogenesis-related secretome of Puccinia helianthi through high-throughput transcriptome analysis. *BMC bioinformatics*, *18*(1), 1-13.
82. Käll, L., Krogh, A. and Sonnhammer, E. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic acids research*, *35*(suppl_2), W429-W432.
83. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27-30.
84. Kaplan, D.T., Thomas, W.K., Frisse, L.M., Sarah, J.L., Stanton, J.M., Speijer, P.R., Marin, D.H. and Opperman, C.H. (2000) Phylogenetic analysis of geographically diverse *Radopholus similis* via rDNA sequence reveals a monomorphic motif. *J. Nematol*. 32, 134– 142.
85. Kimber, M. J., McKinney, S., McMaster, S., Day, T. A., Fleming, C. C. and Maule, A. G. (2007). flp gene disruption in a parasitic nematode reveals motor dysfunction and unusual neuronal sensitivity to RNA interference. *The FASEB Journal,* 21(4), 1233-1243.
86. Klink, V. P., Kim, K. H., Martins, V., MacDonald, M. H., Beard, H. S., Alkharouf, N. W., ... and Matthews, B. F. (2009). A correlation between host-mediated expression of parasite genes as tandem inverted repeats and abrogation of development of female *Heterodera glycines* cyst formation during infection of Glycine max. *Planta*, 230(1), 53-71.
87. Kolde, R. (2012). Pheatmap: pretty heatmaps. *R package version*, *1*(2), 726.
88. Koshel, E. I., Aleshin, V. V., Eroshenko, G. A. and Kutyrev, V. V. (2014). Phylogenetic analysis of entomoparasitic nematodes, potential control agents of flea populations in natural foci of plague. *BioMed Research International*, *2014*.
89. Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., ... and Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS computational biology*, *9*(8), e1003118.
90. Leigh, J. W. and Bryant, D. (2015). popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110-1116.
91. Li, J., Todd, T. C., Oakley, T. R., Lee, J. and Trick, H. N. (2010). Host-derived suppression of nematode reproductive and fitness genes decreases fecundity of *Heterodera glycines* Ichinohe. *Planta*, 232(3), 775-785.
92. Li, J., Xu, C., Yang, S., Chen, C., Tang, S., Wang, J. and Xie, H. (2021). A venom allergen-like protein, rsvap, the first discovered effector protein of *Radopholus similis* that inhibits plant defense and facilitates parasitism. *International Journal of Molecular Sciences*, *22*(9), 4782.
93. Li, X., Zhuo, K., Luo, M., Sun, L. and Liao, J. (2011). Molecular cloning and characterization of a calreticulin cDNA from the pinewood nematode *Bursaphelenchus xylophilus*. *Experimental parasitology*, *128*(2), 121-126.
94. Li, Y., Wang, K., Xie, H., Wang, D. W., Xu, C. L., Huang, X., ... and Li, D. L. (2015). Cathepsin B cysteine proteinase is essential for the development and pathogenesis of the plant parasitic nematode *Radopholus similis*. *International journal of biological sciences*, *11*(9), 1073.
95. Librado, P. and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11), 1451-1452.
96. Lilley, C. J., Goodchild, S. A., Atkinson, H. J. and Urwin, P. E. (2005). Cloning and characterisation of a *Heterodera glycines* aminopeptidase cDNA. *International Journal for Parasitology*, 35(14), 1577-1585.
97. Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *methods*, *25*(4), 402-408.
98. Lourenço-Tessutti, I. T., Souza Junior, J. D. A., Martins-de-Sa, D., Viana, A. A. B., Carneiro, R. M. D. G., Togawa, R. C., ... and Grossi-de-Sa, M. F. (2015). Knock-down of heat-shock protein 90 and isocitrate lyase gene expression reduced root-knot nematode reproduction. *Phytopathology*, *105*(5), 628-637.
99. Lourenço-Tessutti, I. T., Souza Junior, J. D. A., Martins-de-Sa, D., Viana, A. A. B., Carneiro, R. M. D. G., Togawa, R. C., ... and Grossi-de-Sa, M. F. (2015). Knock-down of heat-shock protein 90 and isocitrate lyase gene expression reduced root-knot nematode reproduction. *Phytopathology*, 105(5), 628-637.
100. Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 1-21.
101. Ma, X., Zhan, G., Sleumer, M. C., Chen, S., Liu, W., Zhang, M. Q. and Liu, X. (2016). Analysis of C. elegans muscle transcriptome using trans-splicing-based RNA tagging (SRT). *Nucleic acids research*, *44*(21), e156-e156.
102. MacGowan, J.B. (1977). The burrowing nematode *Radopholus similis* (C, 1893) T 1949. Nematology circulars-Florida Department of Agriculture, 27.
103. Madani, M., Subbotin, S. A., Ward, L. J., Li, X. and De Boer, S. H. (2010). Molecular characterization of Canadian populations of potato cyst nematodes, *Globodera rostochiensis* and *G. pallida* using ribosomal nuclear RNA and cytochrome b genes. *Canadian Journal of Plant Pathology*, *32*(2), 252-263.

104. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., ... and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*, *47*(W1), W636-W641.
105. Maier, T. R., Hewezi, T., Peng, J. and Baum, T. J. (2013). Isolation of whole esophageal gland cells from plant-parasitic nematodes for transcriptome analyses and effector identification. *Molecular Plant-Microbe Interactions*, *26*(1), 31-35.
106. Maizels, R. M., and Yazdanbakhsh, M. (2003). Immune regulation by helminth parasites: cellular and molecular mechanisms. *Nature Reviews Immunology*, *3*(9), 733-744.
107. Mani, V., Reddy, C. S., Lee, S. K., Park, S., Ko, H. R., Kim, D. G. and Hahn, B. S. (2020). Chitin biosynthesis inhibition of *Meloidogyne incognita* by RNAi-mediated gene silencing increases resistance to transgenic tobacco plants. *International journal of molecular sciences*, *21*(18), 6626.
108. Marchler-Bauer, A., and Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Research*, *32*(2), W327-W331. Retrieved from: http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi.
109. Marmonier, A., Perfus-Barbeoch, L., Rancurel, C., Boissinot, S., Favery, B., Demangeat, G. and Brault, V. (2019). In vitro acquisition of specific small interfering RNAs inhibits the expression of some target genes in the plant ectoparasite *Xiphinema index*. *International journal of molecular sciences*, *20*(13), 3266.
110. Möller, S., Croning, M. D. and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, *17*(7), 646-653.
111. Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pages, H., and Gentleman, R. (2009). ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, *25*(19), 2607-2608.
112. Morgan, M., Obenchain, V., Hester, J., and Pagès, H. (2017). SummarizedExperiment: SummarizedExperiment container. *R package version*, *1*(0).
113. Nagaraj, S. H., Gasser, R. B., and Ranganathan, S. (2008a). Needles in the EST haystack: large-scale identification and analysis of excretory-secretory (ES) proteins in parasitic nematodes using expressed sequence tags (ESTs). *PLoS neglected tropical diseases*, *2*(9), e301.
114. Nagaraj, S. H., Gasser, R. B., Nisbet, A. J., and Ranganathan, S. (2008b). In silico analysis of expressed sequence tags from Trichostrongylus vitrinus (Nematoda): comparison of the automated ESTExplorer workflow platform with conventional database searches. In *BMC bioinformatics* (Vol. 9, No. 1, pp. 1-17). BioMed Central.
115. Neuwirth, E., and Neuwirth, M. E. (2011). *Package 'RColorBrewer'*. CRAN 2011-06-17 08: 34: 00. Apache License 2.0.
116. Nicol, P., Gill, R., Fosu-Nyarko, J. and Jones, M. G. (2012). De novo analysis and functional classification of the transcriptome of the root lesion nematode, *Pratylenchus thornei*, after 454 GS FLX sequencing. *International journal for parasitology*, *42*(3), 225-237.
117. Niu, J., Jian, H., Xu, J., Chen, C., Guo, Q., Liu, Q. and Guo, Y. (2012). RNAi silencing of the *Meloidogyne incognita* Rpn7 gene reduces nematode parasitic success. *European Journal of Plant Pathology*, 134(1), 131-144.
118. Niu, J., Liu, P., Liu, Q., Chen, C., Guo, Q., Yin, J., ... and Jian, H. (2016). Msp40 effector of root-knot nematode manipulates plant immunity to facilitate parasitism. *Scientific reports*, 6(1), 1-13.
119. OBannon, J. H. and Taylor, A. L. (1968). Migratory endoparasitic nematodes reared on carrot discs. *Phytopathology*, *58*(3), 385.
120. Pages, H., Carlson, M., Falcon, S., Li, N., Maintainer, M. B. P., DBI, S., and Bimap, S. Q. L. R. A. D. O. A. (2013). Package 'AnnotationDbi'.
121. Papolu, P. K., Gantasala, N. P., Kamaraju, D., Banakar, P., Sreevathsa, R. and Rao, U. (2013). Utility of host delivered RNAi of two FMRF amide like peptides, flp-14 and flp-18, for the management of root knot nematode*, Meloidogyne incognita*. *PLoS One*, 8(11), e80603.
122. Payasi, A., Sanwal, R. and Sanwal, G. G. (2009). Microbial pectate lyases: characterization and enzymological properties. *World Journal of Microbiology and Biotechnology*, *25*(1), 1-14.
123. Peng, H., Cui, J., Long, H., Huang, W., Kong, L., Liu, S., ... and Peng, D. (2016). Novel pectate lyase genes of *Heterodera glycines* play key roles in the early stage of parasitism. *PloS one*, *11*(3), e0149959.
124. Petitot, A. S., Dereeper, A., Agbessi, M., Da Silva, C., Guy, J., Ardisson, M. and Fernandez, D. (2016). Dual RNA-seq reveals *Meloidogyne graminicola* transcriptome and candidate effectors during the interaction with rice plants. *Molecular plant pathology*, *17*(6), 860-874.
125. Petitot, A. S., Dereeper, A., Da Silva, C., Guy, J. and Fernandez, D. (2020). Analyses of the root-knot nematode (*Meloidogyne graminicola*) transcriptome during host infection highlight specific gene expression profiling in resistant rice plants. *Pathogens*, *9*(8), 644.
126. Pierleoni, A., Martelli, P. L. and Casadio, R. (2008). PredGPI: a GPI-anchor predictor. *BMC bioinformatics*, *9*(1), 1-11.
127. ProtComp, S. (2011). 9.0: Predict the sub-cellular localization for plant proteins. *Retrieved May*.
128. Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic acids research*, *33*(suppl_2), W116-W120.

129. Rehman, S., Butterbach, P., Popeijus, H., Overmars, H., Davis, E. L., Jones, J. T., ... and Smant, G. (2009). Identification and characterization of the most abundant cellulases in stylet secretions from *Globodera rostochiensis*. Phytopathology, 99(2), 194-202.
130. Rehman, S., Gupta, V. K., and Goyal, A. K. (2016). Identification and functional analysis of secreted effectors from phytoparasitic nematodes. *BMC microbiology*, *16*(1), 1-18.
131. Rice, P., Longden, I. and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends in genetics*, *16*(6), 276-277.
132. Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139-140.
133. Robinson, M. W., and Connolly, B. (2005). Proteomic analysis of the excretory-secretory proteins of the Trichinella spiralis L1 larva, a nematode parasite of skeletal muscle. *Proteomics*, *5*(17), 4525-4532.
134. Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572-1574
135. Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... and Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, *61*(3), 539-542.
136. Rosso, M. N., Dubrana, M. P., Cimbolini, N., Jaubert, S. and Abad, P. (2005). Application of RNA interference to root-knot nematode genes encoding esophageal gland proteins. *Molecular Plant-Microbe Interactions*, *18*(7), 615-620.
137. Sarkar, D., Sarkar, M. D., and KernSmooth, S. (2015). Package 'lattice'. *Version 0.20*, *33*.
138. Sasser, J.N. and Freckman, D.W. (1987) A World Perspective on Nematology: The Role of the Society. In: Veech, J.A. and Dickson, D.W., Eds., Vistas on Nematology, Hyattsville, Maryland, 7-20.
139. Schlaen, R. G., Mancini, E., Chernomoretz, A., and Yanovsky, M. (2015, October). ASpli: an integrative R package for the analysis of alternative splicing using RNA-seq. In *6th Argentinian Conference on Bioinformatics and Computational Biology*. A2B2C.
140. Shingles, J., Lilley, C. J., Atkinson, H. J. and Urwin, P. E. (2007). *Meloidogyne incognita*: molecular and biochemical characterisation of a cathepsin L cysteine proteinase and the effect on parasitism following RNAi. *Experimental parasitology*, *115*(2), 114-120.
141. Shinya, R., Morisaka, H., Kikuchi, T., Takeuchi, Y., Ueda, M., and Futai, K. (2013). Secretome analysis of the pine wood nematode Bursaphelenchus xylophilus reveals the tangled roots of parasitism and its potential for molecular mimicry. *PloS one*, *8*(6), e67377.
142. Shivakumara, T. N., Papolu, P. K., Dutta, T. K., Kamaraju, D., Chaudhary, S. and Rao, U. (2016). RNAi-induced silencing of an effector confers transcriptional oscillation in another group of effectors in the root-knot nematode, *Meloidogyne incognita*. *Nematology*, *18*(7), 857-870.
143. Sindhu, A. S., Maier, T. R., Mitchum, M. G., Hussey, R. S., Davis, E. L. and Baum, T. J. (2009). Effective and specific in planta RNAi in cyst nematodes: expression interference of four parasitism genes reduces parasitic success. *Journal of experimental botany*, 60(1), 315-324.
144. Smant, G., Stokkermans, J. P., Yan, Y., De Boer, J. M., Baum, T. J., Wang, X., ... and Bakker, J. (1998). Endogenous cellulases in animals: isolation of β-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences*, *95*(9), 4906-4911.
145. Souza, W. D., Carvalho, B. D., and Lopes-Cendes, I. (2018). Rqc: a bioconductor package for quality control of high-throughput sequencing data. *Journal of statistical software*.
146. Steeves, R. M., Todd, T. C., Essig, J. S. and Trick, H. N. (2006). Transgenic soybeans expressing siRNAs specific to a major sperm protein gene suppress *Heterodera glycines* reproduction. *Functional plant biology*, 33(11), 991-999.
147. Swofford, D. L. (2001). PAUP^*: Phylogenetic Analysis Using Parsimony (and other methods) 4.0 b8. *Sinauer, Sunderland, MA*.
148. Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular biology and evolution*, *24*(8), 1596-1599.
149. Tan, J. A. C., Jones, M. G. and Fosu-Nyarko, J. (2013). Gene silencing in root lesion nematodes (*Pratylenchus* spp.) significantly reduces reproduction in a plant host. *Experimental Parasitology*, *133*(2), 166-178.
150. Tan, M., Cobon, J., Aitken, E. and Cook, L. G. (2010). Support for the 'out-of-Southeast Asia' hypothesis for the origin of Australian populations of *Radopholus similis* (Cobb, 1893) (Nematoda: Pratylenchidae). *Systematic parasitology*, *77*(3), 175-183.
151. Tatusov, R. L., Galperin, M. Y., Natale, D. A. and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic acids research*, *28*(1), 33-36.
152. Thorpe, P., Mantelin, S., Cock, P. J., Blok, V. C., Coke, M. C., Eves-van den Akker, S., ... and Jones, J. T. (2014). Genomic characterisation of the effector complement of the potato cyst nematode *Globodera pallida*. *BMC genomics*, *15*(1), 1-15.

153. Tian, B., Li, J., Vodkin, L. O., Todd, T. C., Finer, J. J. and Trick, H. N. (2019). Host-derived gene silencing of parasite fitness genes improves resistance to soybean cyst nematodes in stable transgenic soybean. *Theoretical and Applied Genetics*, *132*(9), 2651-2662.
154. Tyler, L., Bragg, J. N., Wu, J., Yang, X., Tuskan, G. A. and Vogel, J. P. (2010). Annotation and comparative analysis of the glycoside hydrolase genes in *Brachypodium distachyon*. *BMC genomics*, *11*(1), 1-21.
155. Uribe, G. E. M., Bert, W., Vierstraete, A. R., de la Pena, E., Moens, M. and Decraemer, W. (2010). Burrowing nematodes from Colombia and their relationship with *Radopholus similis* populations, *R. arabocoffeae* and *R. duriophilus*. *Nematology*, *12*(4), 619-629.
156. Urwin, P. E., Lilley, C. J. and Atkinson, H. J. (2002). Ingestion of double-stranded RNA by preparasitic juvenile cyst nematodes leads to RNA interference. *Molecular Plant-Microbe Interactions*, *15*(8), 747-752.
157. Urwin, P. E., Lilley, C. J., McPherson, M. J., and Atkinson, H. J. (1997). Resistance to both cyst and root-knot nematodes conferred by transgenic Arabidopsis expressing a modified plant cystatin. *The Plant Journal*, *12*(2), 455-461.
158. Vanholme, B., Van Thuyne, W., Vanhouteghem, K., De Meutter, J. A. N., Cannoot, B. and Gheysen, G. (2007). Molecular characterization and functional importance of pectate lyase secreted by the cyst nematode *Heterodera schachtii*. *Molecular Plant Pathology*, *8*(3), 267-278.
159. Vieira, P., Eves-Van Den Akker, S., Verma, R., Wantoch, S., Eisenback, J. D. and Kamo, K. (2015). The *Pratylenchus penetrans* transcriptome as a source for the development of alternative control strategies: mining for putative genes involved in parasitism and evaluation of in planta RNAi. *PloS one*, *10*(12), e0144674.
160. Vieira, P., Shao, J., Vijayapalani, P., Maier, T. R., Pellegrin, C., Eves-van den Akker, S., ... and Nemchinov, L. G. (2020). A new esophageal gland transcriptome reveals signatures of large scale de novo effector birth in the root lesion nematode *Pratylenchus penetrans*. *BMC genomics*, *21*(1), 1-16.
161. Walawage, S. L., Britton, M. T., Leslie, C. A., Uratsu, S. L., Li, Y. and Dandekar, A. M. (2013). Stacking resistance to crown gall and nematodes in walnut rootstocks. *BMC genomics*, 14(1), 1-13.
162. Wang, C., Lower, S. and Williamson, V. M. (2009). Application of Pluronic gel to the study of root-knot nematode behaviour. *Nematology*, *11*(3), 453-464.
163. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, proteomics* and *bioinformatics*, *8*(1), 77-80.
164. Wang, M., Weiberg, A., Lin, F., Thomma, B.P.H.J. and Huang, H. (2016). Bidirectional cross-kingdom RNAi and fungal uptake of external RNAs confer plant protection. Nat. Plants 2, 16151.
165. Wang, X. R., Cheng, X., Li, Y. D., Zhang, J. A., Zhang, Z. F. and Wu, H. R. (2012). Cloning arginine kinase gene and its RNAi in *Bursaphelenchus xylophilus* causing pine wilt disease. *European journal of plant pathology*, *134*(3), 521-532.
166. Warnes, M. G. R., Bolker, B., Bonebakker, L., Gentleman, R., and Huber, W. (2016). Package 'gplots'. *Various R programming tools for plotting data*.
167. Wilgenbusch, J. C., and Swofford, D. (2003). Inferring evolutionary trees with PAUP. *Current protocols in bioinformatics*, (1), 6-4.
168. Winnenburg, R., Baldwin, T. K., Urban, M., Rawlings, C., Köhler, J. and Hammond-Kosack, K. E. (2006). PHI-base: a new database for pathogen host interactions. *Nucleic acids research*, *34*(suppl_1), D459-D464.
169. Xu, X., Qing, X., Xie, J. L., Yang, F., Peng, Y. L. and Ji, H. L. (2020). Population structure and species delimitation of rice white tip nematode, *Aphelenchoides besseyi* (Nematoda: Aphelenchoididae), in China. *Plant Pathology*, *69*(1), 159-167.
170. Xue, B., Hamamouch, N., Li, C., Huang, G., Hussey, R. S., Baum, T. J. and Davis, E. L. (2013). The 8D05 parasitism gene of *Meloidogyne incognita* is required for successful infection of host roots. *Phytopathology*, 103(2), 175-181.
171. Yadav, B. C., Veluthambi, K., and Subramaniam, K. (2006). Host-generated double stranded RNA induces RNAi in plant-parasitic nematodes and protects the host from infection. *Molecular and biochemical parasitology*, 148(2), 219-222.
172. Yang, Y., Jittayasothorn, Y., Chronis, D., Wang, X., Cousins, P. and Zhong, G. Y. (2013). Molecular characteristics and efficacy of 16D10 siRNAs in inhibiting root-knot nematode infection in transgenic grape hairy roots. *PloS one*, 8(7), e69463.
173. Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research*, *40*(W1), W445-W451.
174. Zhang, C., Xie, H., Xu, C. L., Cheng, X., Li, K. M. and Li, Y. (2012). Differential expression of Rs-eng-1b in two populations of *Radopholus similis* (Tylenchida: Pratylecnchidae) and its relationship to pathogenicity. *European journal of plant pathology*, *133*(4), 899-910.
175. Zhuo, K., Chen, J., Lin, B., Wang, J., Sun, F., Hu, L. and Liao, J. (2017). A novel *Meloidogyne enterolobii* effector MeTCTP promotes parasitism by suppressing programmed cell death in host plants. *Molecular plant pathology*, 18(1), 45-54.

176. Zhuo, K., Chen, J., Lin, B., Wang, J., Sun, F., Hu, L., and Liao, J. (2017). A novel *Meloidogyne enterolobii* effector MeTCTP promotes parasitism by suppressing programmed cell death in host plants. *Molecular plant pathology*, 18(1), 45-54.

PUBLICATIONS FROM THE THESIS

Research publications in peer reviewed journals

- **Blessy MB**, Eapen SJ, Agath M, and Rosana OB (2017). Radobase- A database on burrowing nematodes infesting tropical and sub-tropical crops. *Indian Journal of Nematology* 47(2): 197-200. (NAAS rating: 5.03)
- **Blessy MB and Eapen SJ (2022). Multigene phylogeny of an Indian isolate of *Radopholus similis* infecting black pepper (*Piper nigrum* L.). Indian journal of nematology 52(1): 104-119.**

Research publications in symposiums/seminars

Oral presentation:

- Eapen SJ, Agath M, **Blessy MB** and Rosana OB (2016). RADOBASE- A database on burrowing nematodes infesting tropical and sub-tropical crops. In proceedings of 6th International Conference "Plant, Pathogens and People, Challenges in Plant Pathology to Benefit Humankind" (IPS International Conference 2016), 23-27 February 2016, ICAR, New Delhi, India, pp. 132.
- **Blessy MB**, Krishna PB, Rosana OB and Eapen SJ (2017). MLST analysis of Indian population of burrowing nematode, *Radopholussimilis*, infesting black pepper. In proceedings of National Symposium on "Climate Smart Agriculture for Nematode Management", 11th-13th January, 2017, ICAR- CCARI, Goa, India, pp. 46
- **Blessy MB** and Eapen SJ (2019). "Identifying Novel Target Genes in Burrowing Nematode (*Radopholussimilis* Thorne) – A Transcriptome Based Approach". In XIX International Plant Protection Congress ,10 - 14 November, 2019, Hyderabad, India.

Poster presentation:

- Rosana OB, **Blessy MB** and Eapen SJ (2016). Exposure to salicylic acid and syringaldehyde alters gene expression in burrowing nematode, *Radopholussimilis*. In proceedings of 6th International Conference "Plant, Pathogens and People, Challenges in Plant Pathology to Benefit Humankind" (IPS International Conference 2016), 23-27 February 2016, ICAR, New Delhi, India, pp. 505.
- Rosana OB, **Blessy MB** and Eapen SJ (2017). Exposure to salicylic acid and syringaldehyde alters chemotaxis in burrowing nematode, *Radopholussimilis.* In proceedings of National Symposium on "Climate Smart Agriculture for Nematode Management" 11th-13th January, 2017, ICAR-CCARI, Goa, India, pp. 103and 104
- **Blessy MB**, Sreenath M and Eapen SJ (2018). RNA-seqAnalysis: R based pipeline for completeRNA-seq data analysis. In InCoB-2018, JNU, New Delhi

- **Blessy MB** and Eapen SJ (2019). "Transcriptome and secretome analysis of *Radopholussimilis*, a key nematode pest of tropical crops". In Compendium of National Symposium on “Accelerating Biology towards thinking machines”, 05th-07th February 2019, CDAC, Pune, India, pp. 131-132.
- Swathy KM, Sangeetha T, **Blessy MB**, Anusree VN, Ashok JK, Vijayan KK and Eapen SJ (2019). "How safe are environmental*Pseudomonas aeruginosa* strains -An *in silico* analysis of their whole genomes". In Compendium of National Symposium on “Accelerating Biology towards thinking machines”, 05th-07th February 2019, CDAC, Pune, India, pp. 152-153.

NCBI sequence submission

1. **Blessy MB** and Eapen SJ (2021). Identification of novel target genes from the transcriptome data of *Radopholus simils*. Data accessible at *NCBI,* BioProject accession number **PRJNA698056**
2. **Blessy MB** and Eapen SJ(2019). Identification and characterization of novel genes that may providea safer and more specific control against burrowing nematode, *Radopholus similis*. Data accessible at *NCBI GenBank database*; **MN184219, MN184220, MN184221, MN184222, MN184223, MN184224, MN184225, MN184226, MN184227**.
3. **Blessy MB,**Krishna PB, Rosana OB, Agisha VN and Eapen SJ (2016). Identification of marker genes from burrowing nematode (*Radopholus similis*)from India. Data accessible at *NCBI GenBank database*; **KX378136, KX420702, KX396030, KU530150, KU530151 and KU530152.**

Awards received

1. Received “**Award of Excellence 2018-19 (Best Research Paper Category)**” for the presentation entitled “Transcriptome and secretome analysis of Radopholus similis, a key nematode pest of tropical crops” from ICAR- Indian Institute of Spices

Place: Kozhikode
Date:

Candidate:Blessy M. Baby
MU/EXB/Ph.D./CR.45/Biosc./2015-16/E.13dated 19.01.2016
Reg. No. MU/EXB/Ph.D./CR.54/A.B./2012-13/E.13

Comments by Guide:

Forwarded and recommended by,

Guide: **Dr. Santhosh J. Eapen**
Head and *Principal scientist, Division of Crop protection*
ICAR- Indian Institute of Spices Research, Kozhikode, Kerala

Head of the institution: Director
ICAR-Indian Institute of Spices Research,Kozhikode, Kerala

To
The Registrar (Evaluation)
Mangalore University
Mangalagangothri-574199.

Through,
The Chairman, Department of Biosciences,
Mangalore University, Mangalagangothri-574199